

Impact of Data Privacy Regulations on Recommender Systems Performance

Liben Chen

University of Minnesota, chen7954@umn.edu

Meizi Zhou

Boston University, mzzhou@bu.edu

Yicheng Song

University of Minnesota, ycsong@umn.edu

Gediminas Adomavicius

University of Minnesota, gedas@umn.edu

Abstract

Data privacy regulations empower consumers to control the collection of their personal data. A significant consequence of these regulations is their effect on various data-driven business solutions, especially on personalization technologies and recommender systems. We investigate the potential impacts of diverse real-world data privacy practices (that can be adopted by firms in response to various data protection regulations) on the recommender systems performance. We also examine how these impacts vary across different personalization contexts and applications. In particular, we distinguish between scenarios where the user population exhibits more stable vs. more dynamic (i.e., changing) preferences, as these scenarios often represent distinctly different recommendation settings. We use a simulation framework, carefully seeded with real-world data, for a comprehensive evaluation of the recommender system performance under numerous scenarios, such as different recommendation algorithms, different data privacy practices, different degrees of users' preference dynamics, etc. Extensive computational experiments uncover several robust performance patterns for different data privacy practices and highlight substantial differences between recommendation settings with stable vs. changing user preferences. Furthermore, building upon the comprehensive findings from our computational experiments, we propose flexible, parameterizable data privacy practice designs based on the notion of revelation protection. The proposed approach is designed to provide an effective balance between personalization performance and privacy considerations, as demonstrated using extensive simulation experiments. The findings of this study have significant implications for the design of privacy-aware recommender systems in the context of contemporary data protection regulations. The findings can also be informative to policymakers for understanding the practical implications of various data privacy practices and for designing future policies.

Keywords: Recommender Systems; System Performance; Consumer Privacy; Data Protection Regulations; Stable vs. Dynamic User Preferences; Data Privacy Practices; Simulation-Based Approach.

1 Introduction

The digital economy has experienced substantial growth over the past several decades, facilitated by an online infrastructure that allows for the detailed observation and recording of vast amounts of consumer¹ behavior data. Thus, a digital platform can comprehensively analyze the needs of consumers and provide relevant products and services to them. Recommender systems (RS) represent a class of machine-learning techniques designed to identify and deliver items and content of interest to individuals, leveraging insights into their preferences and behavior (Adomavicius and Tuzhilin 2005, Ricci et al. 2022). Studies have showcased the substantial benefits of RS for users, resulting in an enhanced overall user experience (Konstan et al. 1997, Konstan and Riedl 2012). Specifically, personalization technologies can expedite product discovery, reduce consumers' search costs, and promptly address their immediate needs (Che and Hörner 2018, Li et al. 2022, Yuan et al. 2024), and consequently, have been applied across various industries, significantly boosting revenues in data-driven economies (Gomez-Urbe and Hunt 2015, McKinsey 2024).

The potential of RS heavily relies on the ability to analyze the activity data of individual consumers. Typically such data is collected by tracking users' comprehensive interactions with the service provider, and the data is fed into a machine learning algorithm to infer users' preferences and provide recommendations. However, in recent years, there have been growing concerns about extensive user data collection by companies. For example, a Pew Research Center survey finds that 81% of consumers say the potential risks they face because of data collection by companies outweigh the benefits (Auxier et al. 2019), and an Internet Society survey indicates that 63% of people find connected devices "creepy" in personal data collection (Mori 2019). The public concern over online data practices has garnered significant attention from governments worldwide, leading to a notable trend of regulating such practices. This movement gained momentum in 2018, when the European Union adopted the General Data Protection Regulation (GDPR 2018). GDPR empowers consumers by granting them the right to request the erasure of collected personal data, often referred to as the "right to erasure." Additionally, the ePrivacy Regulation (ePR), also known as the "cookie law," mandates obtaining consumers' consent before using cookies for personal data collection (European Union 2018). The U.S. has not yet enacted a national data protection law, but there are some pioneering regional privacy regulations. Specifically, California Consumer Privacy Act (CCPA), which provides Californians the right to request the deletion of their collected data, went into effect on January 1, 2020 (CCPA 2018a). Virginia Consumer Data Protection Act (VCDPA 2023), which provides Virginia residents much broader opt-out rights than in CCPA, went into effect on January 1, 2023. While this study was ongoing, Colorado, Connecticut, Utah, Washington, and Delaware (CPA 2022, CTDPA 2022, UCPA 2023, WPA 2021, DPDPA 2023) were also at different stages of proposing consumer privacy regulations with similar rights. Data protection laws have been enacted or updated by other nations, including China (CPIPL 2020) and Brazil (GPDPL 2018). Despite some differences in specific legal frameworks, all these laws share a common principle: granting consumers the right to control the collection of data about them.

The implementation of specific data privacy practices (that are made available to consumers) may differ across companies, even as these companies comply with data privacy regulations. Broadly, these data practices

¹ In this paper, we use "consumer" and "user" interchangeably.

are categorized into “delete” and “opt out” options, which pertain to the handling of *past* and *future* data, respectively. The “delete” type of practice allows users to delete past personal data (i.e., already collected by the company) at a given time, but does not exclude the company from collecting future data after the request. In contrast, the “opt out” type of practices can stop the platforms from using future personal data but do not require deleting past data. We provide examples of data privacy practices in Section 3.

Understanding the impact of different data privacy practices on data-driven business activities and, in particular, on RS performance is important for both the consumers’ and the companies’ decision-making. For example, consumers can decide what data privacy actions to request based on how these choices might affect the quality of the recommendations they receive. Companies can deploy various “choice architectures” (Thaler and Sunstein 2009, Lin and Strulov-Shlain 2023), i.e., different data privacy choices (e.g. “delete” or “opt out”) provided to consumers and different ways in which choices can be presented to consumers, when collecting consumer data in various settings, such as the cookie consent pop-ups or the menus for requesting data privacy actions. In this paper, we aim to understand the impacts of personal data privacy practices on RS performance and explore several related issues. Specifically, this study addresses the following research question: How do various common data privacy practices, implementable in accordance with consumer protection regulations, affect the RS performance? It is also crucial to examine how these impacts vary across different personalization contexts and applications. In particular, we distinguish between scenarios where the user population exhibits more stable vs. more dynamic preferences, as these scenarios often represent distinctly different recommendation contexts.² Consequently, even the same type of data privacy practice may have different impacts on the RS performance in different recommendation contexts. Furthermore, we explore how the effects of different data privacy practices are moderated by several factors, including the specific recommendation algorithms used, the size of the privacy-conscious user segment, and the extent of users’ reliance on RS. Moreover, we explore how some specialized machine learning strategies that respond to limited data usage, such as *incremental learning*, react to various data privacy practices. Finally, we go beyond existing choices for privacy practices by proposing and evaluating parameterizable data privacy practice designs that can help companies and users navigate the privacy-personalization tradeoff effectively.

To investigate these issues, we adopt a simulation approach that is seeded using real-world e-commerce data. For complex research questions encompassing multiple dimensions, a simulation-based methodology presents several advantages over alternatives. The simulation-based approach enables experiments on large-scale user populations that have extensive interactions with an RS across hundreds of iterations while allowing for targeted manipulations and precise control (and separating the effects) of different factors, e.g., data privacy policies, users’ preference dynamics, privacy-conscious population size, recommendation algorithms, etc. In contrast, conducting real-world large-scale field experiments can often prove challenging in terms of controlling key factors; for example, it may not be feasible to artificially induce certain desired levels of preference dynamics in a real-world consumer population, and implementing certain recommendation algorithms and certain data privacy practices may lead to substantial decrease in recommendation quality for consumers and in revenue losses for companies. Also, unlike analytical modeling approaches that

²For instance, streaming platforms such as Netflix and Spotify, where user preferences largely reflect *inherent tastes* for different types of content, may exhibit more stable, longer-term user preferences. Conversely, e-commerce platforms like Amazon and Best Buy, where preferences are often influenced by *immediate needs*, may demonstrate more dynamic user preferences.

typically require highly simplified assumptions about the outcomes of complex (often black-box) machine learning algorithms, simulation testbeds offer high configurability and can directly integrate and test actual recommendation algorithms. Therefore, several previous studies (D’Amour et al. 2020, Zhang et al. 2020, Zhou et al. 2023) on longitudinal effects of user-RS interactions have used a simulation-based approach.

This study provides a number of contributions. In particular, we introduce a simulation framework for exploring the longitudinal impacts of data privacy practices on the RS performance, as discussed in Section 4. The framework extends existing work by Zhang et al. (2020) on explicit user feedback scenarios, i.e., where users express their “ground truth” preferences to RS explicitly as ratings, into implicit user feedback settings, where the users’ preferences are inferred from user interactions, e.g., clicks or purchases. The proposed framework facilitates the examination of how various data privacy practices impact RS performance and enables to study this in a thorough and nuanced manner. E.g., the simulation framework also introduces a versatile approach to modeling the dynamic nature of users’ preferences, enabling the simulation of individual preference changes. Differentiating between settings with stable vs. dynamic user preferences is highly informative, as these settings can represent different types of RS application domains. We leverage the simulation framework and obtain numerous robust insights into the impact of data privacy practices on RS performance. As one key example, we find that existing data privacy practices may have highly differing impacts on the RS performance in different recommendation settings. Specifically, when users have stable or minimally changing preferences, opting out of future data collection may have little to no effect on RS performance, whereas the deletion of past data can drastically reduce performance. In contrast, when users have significantly changing preferences, opting out of data collection can severely impair RS performance, whereas requests for data deletion can even enhance the system’s effectiveness. Many additional findings and insights – related to robustness of RS performance patterns across different recommendation algorithms, the impact of privacy-conscious population size, the behavior of *incremental learning* under data privacy practices, etc. – are discussed throughout Section 5. Moreover, Section 6 introduces additional data privacy practice designs based on the notion of revelation protection, which allows platforms and consumers to parameterize data actions in a more granular manner, so that the more limited, partial data deletions (or opt outs) can still provide privacy benefits while retaining personalization performance as much as possible. The effectiveness of the proposed designs is demonstrated using a set of computational experiments. Our findings have important practical implications and future directions, as discussed in Section 7.

2 Related Work

Practical Impacts of Data Privacy Regulations. This study joins the scholarly discussion on the practical impacts of data privacy/protection regulations. In particular, a number of studies have investigated the impact of data regulation policies on targeted advertising (Goldfarb and Tucker 2011, Goldberg et al. 2019), e-commerce (Sun et al. 2020), innovation (Jia et al. 2018), and healthcare (Adjerid et al. 2016). Most of them focus on the economic outcomes of such policy changes. For example, Goldfarb and Tucker (2011) show that restricting online tracking techniques in the European Union leads to a substantial decline in the effectiveness of online advertising. As another example, Sun et al. (2020) set up a field experiment on an e-commerce platform to study the customer reactions to banning personal data usage by the platform’s RS;

they find that such change can lead to a sharp decrease in customer engagement and transactions. Our study focuses on the algorithmic RS performance implications of data privacy regulations, which is an under-studied topic and provides a complementary perspective to the prior literature that has focused on the economic outcomes. Furthermore, many previous studies focus on empirical impacts of a specific regulation or practice, often within only a specific locale or on a specific platform (Sun et al. 2020, Adjerid et al. 2016, Goldfarb and Tucker 2011). Yet, the data practices reflected across different data privacy regulations can be quite diverse. Therefore, in this study, we take a more general perspective, first, by providing a broad picture of different consumer privacy choices under a wide variety of privacy regulations in Section 3 and, second, by computationally comparing the personalization performance impacts of different data practices to provide a more general understanding applicable more widely across the current regulatory landscape.

Consumer Privacy Choices. This research also connects to an important stream of IS literature on the consumer privacy choices. Consumer privacy choices are often complicated, context-dependent, and subject to the user's privacy valuation (Acquisti et al. 2016). Previous research suggests that users may exhibit highly varying privacy valuations, which affects the likelihood of these users enacting privacy practices (Li 2014, Bansal et al. 2010). Also, IS literature has thoroughly documented the mismatch between stated privacy concerns and actual disclosure behaviors (Alashoor et al. 2023, Kokolakis 2017). The reason may be that privacy-related decision making is often affected by incomplete information, bounded rationality, and psychological biases (Acquisti and Grossklags 2005, Alashoor et al. 2023). Another important aspect that influences the users' privacy choices is the privacy choice architecture (e.g., the availability of choices, the setting of default choice, and the presence of monetary incentives). Gross and Acquisti (2005) find that few users (1.2%) alter the permissive default privacy settings on social media. Athey et al. (2017) show that both small incentives and small costs generate large shifts in a consumer's supply of private information. In contrast to the literature which focuses on understanding consumer behavior and decision making when facing privacy choices, our work contributes to the design of the choice architecture for privacy policies by revealing the impacts of different existing data privacy choices on RS, and by exploring new designs for data privacy practices. Specifically, we introduce a parameterizable design for data privacy practices that allows for a more nuanced exploration of the tradeoff between the RS performance and personal data revelation.

Data Privacy in RS. Our work builds upon previous research on privacy-related issues in RS (Friedman et al. 2015, Kobsa 2007). For example, the famous Netflix Prize (Bennett et al. 2007) not only sparked the algorithmic advancements for RS, but also highlighted the privacy protection challenges. Narayanan and Shmatikov (2008) demonstrated the challenges of ensuring anonymity in transaction and preference records typical to RS applications. The sparseness of these extensive datasets means that individual records often lack "similar" counterparts, making it possible to single them out and re-identify them with relatively little background information; e.g., given public IMDb user ratings, the adversary could successfully identify some of the Netflix users (Narayanan and Shmatikov 2008). Existing privacy protection techniques typically focus on specific application contexts or privacy threats. For example, *differential privacy* (i.e., adding a controlled amount of noise to the real data) (Chen et al. 2021, Khaliq et al. 2022) focuses on privacy protection when de-identified data is shared publicly; and *federated learning* (i.e., distributed machine learning from different data sources without directly sharing training data) (Li et al. 2020, Minto et al. 2021, Wang et al. 2022)

facilitates privacy protection when building a global predictive model across multiple stakeholders. Many of these approaches are not directly applicable for providing the specific capabilities codified in the data privacy regulations, which reflect consumer concerns about (and, thus, the need for a direct control of) the amount of personal data available to online platforms. Thus, additional approaches should be adopted and/or developed for addressing the specific challenges raised by data privacy regulations. As part of this study, in addition to introducing a parameterizable design for data privacy practices (as mentioned earlier), we also investigate the applicability of the *incremental learning* approach, which allows the platform's RS to retain the *inferences* (e.g., latent preference representations) about the users rather than raw user data.

User Preference Dynamism and RS. Finally, our study contributes towards understanding how user preference dynamics influence RS performance. Previous studies have shown that user preferences may be subject to change in the process of user-system interactions, and the extent of the change depends on various factors, such as the application domain and user characteristics (Rafailidis and Nanopoulos 2015, Jannach et al. 2017, Evans and Kasirzadeh 2021). As one example, user preferences may change endogenously in the process of consuming new items – the consumption experience of specific items alters the user's prior opinion about the producer of items (Rafailidis and Nanopoulos 2015), and the user may dislike or like more items from the same producers. As another example, e-commerce users visit the site with a specific intent for purchase, and this intent may change from one visit to another (Moe 2003). Studies have shown that the preferences of e-commerce users are often session-based and highly dynamic (Jannach et al. 2017). Previous studies provide evidence that modeling these user preference dynamics is vital to improving both system performance and consumer welfare (Koren 2009, Kang and McAuley 2018, Song et al. 2019). We contribute to this field of study by examining the interplay between dynamic user preferences and privacy regulations, and its influence on the RS performance, a topic not yet addressed in the existing literature.

3 Personal Data Protection Practices

Data privacy/protection regulations empower consumers to exert control over their personal data collected by companies. These regulations require that affected companies offer choices to consumers concerning the management of their personal information (CCPA 2018b). Table 1 provides a broad overview of the common data privacy practices via example regulations and company offerings.

Regarding consumer rights to manage their data, a common provision within data protection regulations is the ability to “delete” previously collected information, i.e., delete past data. The implementation of this provision varies among companies. For instance, Amazon requires the complete closure of a user's account for data erasure (Amazon 2023), while eBay permits users to delete their data without terminating their account (Ebay 2023). Customers of Best Buy can submit a form or call customer service requesting deletion, as shown by Figure A11 in Appendix A. To ensure that personal data is not utilized, users must submit periodic deletion requests, a process that some companies explicitly accommodate. E.g., YouTube offers the option to automatically erase historical data that is older than a certain user-specified time period,³ and Etsy enables consumers to appoint an authorized agent to submit requests on their behalf (Etsy 2023). Apple's ITP

³ See the “auto-delete” option on myactivity.google.com, and an example is shown in Appendix A.

Table 1: Overview of Data Privacy Regulations, Personal Data Definitions, and Data Privacy Practices

A. Example Data Privacy Regulations and Company Policies			
Abbreviation	Full Title	Jurisdiction/Company	Reference
CCPA	California Consumer Privacy Act	California, the U.S.	CCPA (2018b)
CTDPA	Connecticut Data Privacy Act	Connecticut, the U.S.	CTDPA (2022)
CPA	Colorado Privacy Act	Colorado, the U.S.	CPA (2022)
DPDPA	Delaware Personal Data Privacy Act	Delaware, the U.S.	DPDPA (2023)
GDPR	General Data Protection Regulation	European Union	GDPR (2018)
UCPA	Utah Consumer Privacy Act	Utah, the U.S.	UCPA (2023)
VCDPA	Virginia Consumer Data Protection Act	Virginia, the U.S.	VCDPA (2023)
WPA	Washington Privacy Act	Washington, the U.S.	WPA (2021)
ATT	App Tracking Transparency	Apple	Apple (2019)
ITP	Intelligent Tracking Prevention	Apple	Adjust (2021)
ePR	ePrivacy Regulation	European Union	European Union (2018)
B. Personal Data Definitions			
Regulation	Definition of Personal Data		
GDPR	“Personal data” means “any information relating to an identified or identifiable natural person”.		
CCPA	“Personal information” means “information that identifies ... a particular consumer or household. Personal information includes ... inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer’s preferences ...”		
VCDPA, CTDPA, WPA, DPDPA, UCPA, CPA	“Personal data” means any information that is linkable to an identified or identifiable natural person. “Personal data” does not include de-identified data or publicly available information.		
C. Data Privacy Practices			
Type	Example Regulations		Example Company Offerings
Delete past data	- GDPR “Art. 17 ... The data subject shall have the right to obtain from the controller the erasure of personal data ... ” - CCPA “1798.105. ... A consumer shall have the right to request that a business delete any personal information ...” - VCDPA “59.1-577 ... consumer request to exercise the right ... to delete personal data ...” - CPA “6-1-1306 ... A consumer has the right to delete personal data ...” - CTDPA “Sec 4 ... A consumer shall have the right to ... delete personal data ...” - UCPA “13-61-201 ... A consumer has the right to delete the consumer’s personal data ...” - WPA “Sec 6 ... A consumer has the right to delete personal data concerning the consumer ...” - DPDPA “12D-104. ... delete personal data provided by, or obtained about, the consumer...”		- Amazon requires the complete closure of a user’s account for data erasure (Amazon 2023). - eBay permits users to delete their data without terminating their account (Ebay 2023). - YouTube offers the option to automatically erase history after a certain time period (Appendix A Figure A12). - Etsy enables consumers to appoint an authorized agent to submit requests on their behalf (Etsy 2023). - Apple’s ITP limits first-party cookies’ lifetime to 7 days (Adjust 2021).
	- ePR “Art. 8 ... the collection of information from end-users’ terminal equipment ... shall be prohibited, except on the following grounds: ... the end-user has given his or her consent ...” - VCDPA “59.1-577 ... consumer request to exercise the right ... to opt out of... profiling in furtherance of decisions ...” CPA “6-1-1306 ... A consumer has the right to opt out of ... profiling in furtherance of decisions ...” - CTDPA “Sec 4 ... A consumer shall have the right to opt out of ... profiling in furtherance of solely automated decisions ...” - WPA “Sec 6 ... A consumer has the right to opt out of ... profiling in furtherance of decisions ...” - DPDPA “Opt out of the processing of the personal data for ... profiling in furtherance of solely automated decisions ...”		- Websites covered by GDPR need to include cookie consent banner or cookie wall (Appendix A Figure A13). - Apple’s ATT policy allows consumers to opt out of data collection through options like the “ask app not to track” (Apple 2019).

Note: The opt-out rules in GDPR, CCPA, VCDPA, CPA, CTDPA, UCPA, WPA, DPDPA also include opting out of the selling of personal data and targeted advertising, which are not related to recommender system performance, so we omit them from the table.

mandates that Safari browser clears the cache of customer data after seven days (Apple 2019). Importantly, companies are not obligated to stop future data collection for a user following a deletion request. Based on these observations, we explore two typical data practice scenarios: a one-time data deletion and data deletion with high frequency. Another common provision within data privacy regulations gives consumers the right to opt out of the future data collection, e.g., rejecting the use of cookies via a cookie consent banner required by “cookie law” (see an example in Appendix A) or opting out of data profiling as codified by the VCDPA. Privacy policies from companies, such as Apple’s ATT, provide mechanisms for consumers to opt out of data collection through options like the “ask app not to track” (Adjust 2021). When consumers choose to opt out of data collection or usage, this prevents RS from incorporating their new data into predictive models. Accordingly, this study incorporates an analysis of such data practices.

The proportion of users engaging in specific data privacy practices varies across platforms and over time. In the U.S., while only consumers from certain states are legally entitled to “delete” their past data, the frequency of deletion requests has increased markedly. For example, there has been a 34-fold increase in such requests at Best Buy and a 5-fold increase at Amazon between 2020 and 2022 (BestBuy 2023, Amazon 2023). In 2021, around 62% of iPhone owners chose to opt out of future data collection (Leswing 2021), and only around 5%-16% of users permitted mobile apps to track their data (Laziuk 2021). Reflecting this variation, our simulation allows to adjust the ratio of the *affected* users, i.e., users who exercise certain data practices. This enables us to examine the consequences of these practices not only on the users who choose to exercise their data privacy rights but also on those who continue to allow personal data collection.

Beyond the commonly accepted categories of personal data, such as demographic information and user activities, the classification of user-related *inferences* (e.g., learned user models or representations) as personal data is subject to debate across different regulations. In Table 1.B, we provide a comparison of how the definition of personal data varies across regulatory frameworks. Some regulations, like VCDPA, define personal data mainly as identification information (e.g., online identifiers such as IP and device ID) and do not include de-identified data as personal data. Meanwhile, CCPA defines personal data as not only identification data but also online activities (e.g., viewed and searched products or content, length of visits to certain pages, and page interaction information including scrolling, clicks, and mouse-overs), transactions (e.g., purchase history), and learned user inferences (e.g., derived interests or preferences from a model trained on personal data) (BestBuy 2023). The learned inferences could be user profiles (e.g., Google Ads personalization describes a user with 100+ keywords) or implicit representations of users (e.g., model’s weights/embeddings representing user features in RS models). Although these inferences do not directly expose the user’s sensitive personal information, they are still associated with privacy concerns, e.g., inference attacks (Hu et al. 2022), where an attacker can infer the user’s sensitive personal information from the learned inferences that are leaked. Specifically, studies have found that learned embeddings of AI models can be reverse-engineered to reveal sensitive attributes of the dataset used to train these models (Song and Raghunathan 2020). However, some of the current privacy regulations do not explicitly designate learned inferences as a category of personal data, leaving their classification ambiguous. Consequently, the debate remains unresolved as to whether the practical advantages of preserving learned inferences outweigh the associated privacy risks. Thus, for the situations where the learned inferences might not fall under personal data privacy protections, this study

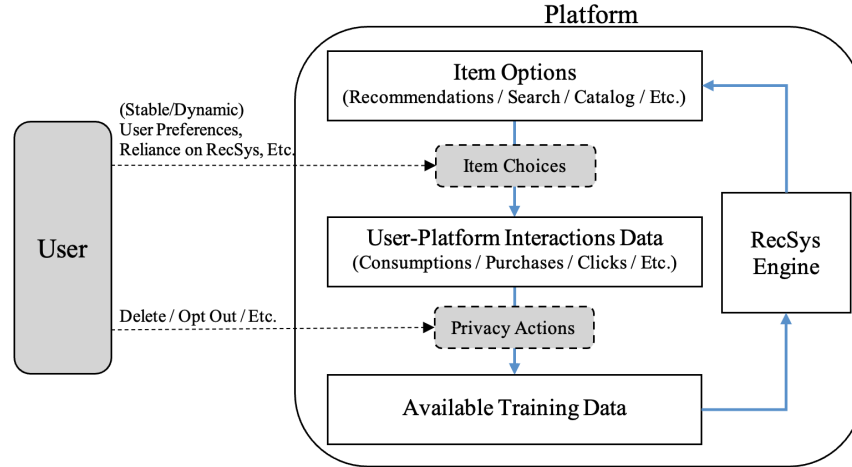


Figure 1: Conceptual diagram: key components of user-platform interaction setting explored in this study

investigates the benefits of the incremental learning approach that utilizes these learned inferences in an attempt to improve the RS performance under data privacy practices.

4 Simulation Framework

To systematically explore the impact of different data privacy practices on the RS performance, we adopt and enhance a simulation framework from prior literature (Zhang et al. 2020). The framework models a generic feedback loop between the user and the RS engine on the platform, as shown in Figure 1. In this feedback loop, RS recommends items; the user chooses items to consume from different options (through recommendations or other capabilities provided by the platform, such as search or catalog browsing); the platform logs users' interaction data; the users enact privacy actions on the logged data, on which RS is subsequently (re)trained. This framework also manipulates (and controls for) various factors, such as user preference dynamics, recommendation algorithms, data privacy practices, and ratio of privacy-conscious (i.e., affected) users, over an extended period through a user-system feedback loop. We enhance the previous simulation framework (Zhang et al. 2020) in several key ways. First, we incorporate modeling of user data privacy practices, considering factors such as the nature of these practices, timing of user requests, and the percentage of users making such requests. Second, whereas the original framework assumes stable user preferences, our approach allows to model the situations where user preferences can change over time. This enables us to study how a change in user preferences interacts with data privacy practices, offering a more nuanced understanding of its impact on RS. Finally, we extend the original framework from explicit-feedback RS, which collect users' preferences for items as explicit numeric ratings, to implicit-feedback RS. I.e., our study encompasses scenarios where users' preferences are inferred via implicit user feedback, such as clicks, purchases, or consumptions. Such scenarios are increasingly prevalent on many online platforms, especially in contexts like e-commerce and media streaming platforms.

4.1 General Simulation Procedure

From a high-level perspective, our simulation approach consists of two main phases. In the first (i.e., pre-simulation) phase, we properly initialize the simulation testbed by seeding it with realistic user and item populations and by integrating a specific recommendation algorithm. The user and item populations are represented by user preference profiles and item content profiles consisting of latent representations (embeddings) learned from real-world datasets by machine learning techniques. Then, in the second (i.e., main) phase, the simulation testbed uses user and item populations as well as the recommendation algorithm to generate recommendations potentially leading to users' consumption choices (based on recommendations and user preferences). As a result, a comprehensive set of user-item consumption data is collected over time, which represents personal user data that can be employed for further improvement of the recommendation algorithm. Because the main goal of this study is to understand the impact of data privacy practices on RS performance, the second phase can be viewed as consisting of two sub-phases – before the data privacy action is requested and after. Data privacy practices reduce the RS access to (some portion of) personal user data, consequently influencing the recommendations, and users react accordingly to these changes. By examining the impact of various personal data privacy practices on RS performance, we can comprehensively analyze their effects. A more detailed overview of the simulation procedure is provided in Algorithm 1, and we discuss key aspects below. Appendix B lists the main simulation components and important variables.

Learning Seed User/Item Embeddings. To generate realistic user and item populations for our experiments, we learn latent vector representations for each user and item from a representative real-world dataset. We employ the Logistic Matrix Factorization (LMF) model (Johnson 2014) to decompose the real-world user-item interaction matrix into two low-rank sub-matrices. Each user $u \in U$ and each item $i \in I$ is then represented by an embedding, which is a M -dimensional vector of latent features (in our experiments, $M = 20$). Consequently, item i is represented by vector $e_i \in \mathbb{R}^M$, and similarly user u is represented by vector $e_u \in \mathbb{R}^M$. Hyperparameters of the LMF model are fine-tuned via grid search for accurate learning.

Computing User Preferences for Items. The latent user/item representations learned from the real-world dataset are then used by the simulation testbed to model users' actual ("ground truth") preferences for items. Specifically, each user u 's preference for item i is simulated by $p_{ui} = e_u^t e_i$. Also, for each user u , let $PrefList_u$ represent a list of items ordered by user u 's preference, i.e.:

$$PrefList_u = \{i_1, i_2, \dots, i_{|I|}\}, \text{ where } p_{ui_1} \geq p_{ui_2} \geq \dots p_{ui_{|I|}} \quad (1)$$

Computing System Recommendations. Recommender systems regularly update their models to incorporate recent observations and maintain optimal performance. When the RS model is trained during the simulation, we use a single recommendation algorithm for the entire population and train the model with the use of all available historical data. Since re-training the model with the full historical dataset can be time-consuming, we perform model re-training on a *weekly* basis in our simulations to balance computational efficiency and recommendation effectiveness. After each re-training, the algorithm predicts consumption probability $pred_{ui}^t$ of each user-item pair and generates a recommendation list $RecList_u^t$ for each user:

$$RecList_u^t = \{i_1, i_2, \dots, i_{|I|}\}, \text{ where } pred_{ui_1}^t \geq pred_{ui_2}^t \geq \dots pred_{ui_{|I|}}^t \quad (2)$$

Algorithm 1 General Simulation Procedure

```

1: procedure [INITIALIZATION ( $t = 0$ )]
   [Step 0: System Initialization]
2:   Initialize item content  $e_i$  and user preference  $e_u$ , and generate  $PrefList_u$  for each user  $u$ 
3:   Initialize  $ConsData_u^0$  for each user  $u$  by letting  $u$  consume for  $T_{init}$  periods based on  $PrefList_u$  without recommendation
4:   Let  $TrainData_u^0 = ConsData_u^0$  for each user (i.e., all initial consumptions are available to RS as training data)
5:   Prepare the list of recommendations for each user  $u$  for the first time period, i.e.,  $RecList_u^1 = GenerateRecs(TrainData_u^0)$ 
6: end procedure
7: procedure [MAIN SIMULATION ( $t = 1, 2, 3, \dots$ )]
   [Step 1: Update User Preference Change, if/when needed]
8:   If  $t = T_{PrefChange}$ , assign each user a new preference profiles  $\hat{e}_u$ 
   [Step 2: Main Interaction/Consumption/Feedback Process]
9:   For each user  $u$  who consumes at time  $t$  do
10:    User  $u$  chooses to consume item  $i$  from either  $PrefList_u$  or  $RecList_u^t$ , according to  $u$ 's degree of reliance on RS
11:    Update consumption data, i.e.,  $ConsData_u^t = ConsData_u^{t-1} \cup i$ 
12:    Update training data, i.e.,  $TrainData_u^t = TrainData_u^{t-1} \cup i$ 
13:    Determine the time of the next consumption for  $u$ , i.e.,  $UserNextCons_u = UserNextCons_u + \Delta$ 
14:   end for
   [Step 3: System Update and Evaluation]
15:   Evaluate the recommender system performance
16:   Prepare the list of recommendations for each user  $u$  for next time period, i.e.,  $RecList_u^{t+1} = GenerateRecs(TrainData_u^t)$ 
   [Step 4: Apply Data Privacy Practices, if/when needed]
17:   If  $t \geq T_{DataReq}$ , exercise data practices  $DataRequest_u$  for each user  $u$  and obtain  $TrainData_u^t$  from  $ConsData_u^t$  accordingly
18: end procedure

```

Simulating User-Item Interactions/Consumptions. Each temporal iteration t of our main simulation represents a separate day, during which some subset of users may choose to visit the platform and consume an item. With respect to which items get chosen by the users for consumption when they visit the platform, our simulation incorporates the following two key aspects: (a) users' own inherent preferences/needs and (b) observed recommendations from RS. Specifically, we implement a two-urn choice model (Fleder and Hosanagar 2009) for selecting items to be consumed. List $PrefList_u$, defined in Equation (1), represents the first urn, from which users make choices based on their preferences/needs. List $RecList_u^t$, defined in Equation (2), represents the second urn, from which users make choices by relying on RS. For each individual consumption, a user follows either $RecList_u^t$ with probability w , or $PrefList_u$ with probability $(1 - w)$. Here, w represents the degree of a user's *reliance* on RS. In most simulation experiments, w is set to 20% by default, but we also provide some explorations with varying values of w . We also assume that we are not in a "repeated consumption" application domain, i.e., that the users do not purchase or consume the same item twice. Thus, we exclude the previously consumed items from both $RecList_u^t$ and $PrefList_u$ before a user makes consumption decision. Furthermore, we assume that items have unlimited inventory (e.g., as is the case with most digital goods), allowing simultaneous consumption by multiple users.

Within each urn, the user follows the exponential-decay function (Equation (3)) to choose an item from the list. This function represents the realistic user behavior. For example, users' clicks on Google search results follow an exponential decay pattern based on the rankings with most clicks focusing on the top-5 results,⁴ and most users interact with lists by scanning them from top to bottom and clicking on the first item that seems relevant, as opposed to holistically evaluating the entire list (Klöckner et al. 2004, Collins et al. 2018). The use of exponential-decay function is also consistent with prior work on using simulation-based approaches in RS (Zhang et al. 2020). Specifically, the probability of interacting with an item at rank r is

⁴<https://backlinko.com/google-ctr-stats>

$$prob_r = (\beta - 1) \cdot \beta^{-r} \quad (3)$$

where β ($\beta > 1$) controls how far down the list users are likely to explore before choosing an item. The larger β values, the more user choices are concentrated on the top-ranked items. In our experiments, $\beta = 1.3$.⁵

After each simulated interaction (i.e., item consumption) by user u , we use a normal distribution to model the next interaction to occur after time gap $\Delta_{ut} \sim N(\mu_u, \sigma_u)$, where μ_u, σ_u are the user-specific mean and variance of the time gap between interactions from user u estimated from the real-world dataset.⁶

Evaluating Recommendation Performance. The key goal of our study is to assess the impact of data privacy practices on the predictive (i.e., accuracy) performance of RS. For our main experiments, we use the “hit rate” ($Hit@K$), a common accuracy metric for top- K recommendation lists. At each time t , $Hit@K_t$ is the fraction of users for whom the consumed item was a part of top- K recommendations, calculated as:

$$Hit@K_t = |UHit_t^K| / |UConsume_t| \quad (4)$$

where $|UHit_t^K|$ is the number of users for whom the consumed item is among the top K recommendations, and $|UConsume_t|$ is the total number of users who consumed at time t . In our experiments, $K = 20$.

For completeness, we have also explored additional performance metrics, including several ranking-based accuracy metrics, such as $NDCG@20$ and $MRR@20$ (see Appendix D for details), as well as item discoverability metrics, such as recommendation diversity (Appendix W) and longtailness (Appendix X).

Incorporating Data Privacy Practices. There are several aspects to modeling the impact of data privacy practices on RS performance. First, in order to understand the *temporal* effect of data privacy regulations, in our simulations we model the privacy regulation “launch” date as $T_{DataReq}$. Before $T_{DataReq}$, the activity data of all users is available to the RS. After $T_{DataReq}$, various data privacy practices become available, and access to user activity data is subsequently restricted based on the privacy requests enacted by users.

Second, although privacy regulations provide users with rights to control their personal data collection by online platforms, the proportion of users acting upon these rights may vary. Thus, in order to understand the *population* effect, we model that a portion of users (referred to as “affected” users) will choose to make the data privacy requests as soon as they are available, at $T_{DataReq}$. The remaining users (referred to as “unaffected” users) will not be affected by the regulations, and their data will continue to be collected and fully available to the RS. In most simulation experiments, we set the portion of affected users to be 50% by default, but we also provide some explorations for both larger and smaller affected user populations.

Third, to understand the effects of *specific data privacy practices*, we explore four key data practices, denoted as $DataRequest_u$ in Algorithm 1, based on empirically observed practices described in Section 3:

- **NO_DELETE:** The affected users make no data privacy requests at $T_{DataReq}$, and the system keeps collecting their data. In other words, all historical data of all users is available to RS at all times.

⁵E.g., when $\beta = 1.3$, the probability of selecting top-1 item is 23.1%, one of top-3 items 54.5%, and one of top-10 items 92.7%. For robustness, we provide the main experiment results with $\beta = 1.1, 1.3$, and 1.7 in Appendix C; the results are highly consistent.

⁶For reference, the average values (i.e., across all users) of μ_u and σ_u in our experiments are 28 and 15 days, respectively.

- **ONE_TIME**: The affected users delete their historical data just once at $T_{DataReq}$; their new data generated after $T_{DataReq}$ can still be collected and is available for the recommender engine.
- **HIGH_FREQ**: The affected users delete their historical data periodically (every week), starting at $T_{DataReq}$; only their newest data (i.e., generated after the previous deletion) is available to the system.
- **OPT_OUT**: The affected users opt out of future data collection at $T_{DataReq}$; i.e., the RS cannot use any affected users' data generated after $T_{DataReq}$, only data from before $T_{DataReq}$ is available.

Incorporating User Preference Change. In some application domains, user preferences may shift based on the users' current needs and circumstances. In our simulation approach, we introduce the ability to model user preference changes. In settings with stable user preferences, each user is represented by the original user embedding e_u , and this representation remains constant throughout the simulation process. In dynamic preference settings, we allow for the user embedding to vary over time. To model the changing preferences of user u in a realistic manner, we randomly select another existing user v and use e_v as a realistic candidate for u 's "new" preferences $\hat{e}_u$. More specifically, to model a gradual change of user preferences, we vary the updated representation $\hat{e}_u$ between embeddings e_u and e_v (both of which were seeded from a real-world dataset) by introducing parameter α (where $0 \leq \alpha \leq 1$) as follows:

$$\hat{e}_u = (1 - \alpha) \times e_u + \alpha \times e_v, \quad u, v \in U \text{ and } u \neq v \quad (5)$$

When $\alpha = 0$, u 's embedding remains identical to the original ($\hat{e}_u = e_u$); when $\alpha = 1$, u 's embedding is reflecting completely different preferences ($\hat{e}_u = e_v$); and, the higher the α , the larger the preference change.

With the updated user representation $\hat{e}_u$, user u 's $PrefList_u$ is reconstructed accordingly, by re-estimating $p_{ui} = \hat{e}_u e_i$ based on Equation (1). Correspondingly, user u will use the updated $PrefList_u$ for making the subsequent item choice decisions.

In order to identify the impact of different degrees of the user preference change in a clear and controlled manner, we use homogeneous user populations in our simulation experiments, where the entire population has the same degree of preference change α . Also, modeling a one-time preference change (at pre-specified time $T_{PrefChange}$) has been sufficient for demonstrating key insights in our study. However, it is important to point out that the proposed simulation approach can readily accommodate settings with heterogeneous preference change levels and with multiple changes to user preferences over time.

4.2 Empirical Setting

In our main experiments, we seed the simulation with a real-world dataset from a major multi-channel retailer in the U.S. The retailer has both an online e-commerce platform and a large number of physical stores. The dataset contains order data from June 1, 2018 to July 1, 2019 for a randomly selected group of users. In total, the dataset has 18,938 users, 40,716 items, and 110,953 orders (i.e., both online and offline user-item interactions).

In addition, we repeated our experiments on two other public datasets, representing different application domains – *Amazon Book* (book purchases) and *Retailrocket* (product views) – with highly consistent results

regarding the impact of data privacy practices on RS performance.⁷

We first apply Logistic Matrix Factorization (Johnson 2014) on the real-world data to derive vectors of latent user preferences and latent item content. Then we generate the historical user-item interactions by letting users consume organically, i.e., without recommendations, for period $T_{init} = 365$ days. During this period, we simulate 115,535 user-item interactions used as historical data $ConsData_u^0$ – roughly equivalent to the activity volume of our real-world dataset. After T_{init} , we begin the main simulation period, which lasts for 65 weeks (i.e., 455 days). During this period, RS is (re)trained every week and evaluated using $Hit@20$. In the first 24 weeks, there are no data privacy regulations, and RS can reach a relatively stable, steady-state performance levels. The data privacy shock happens on $T_{DataReq}$, i.e., at the end of 24th week. After $T_{DataReq}$, we observe the longitudinal impacts of different data privacy practices on RS performance.

We implement six popular implicit-feedback RS techniques from diverse algorithmic families, including: two matrix factorization methods, Alternating Least Squares (ALS) and Bayesian Personalized Ranking (BPR) (Hu et al. 2008, Rendle et al. 2012); an item-based nearest-neighbor method (ItemKNN) (Sarwar et al. 2001); and two deep learning methods, Neural Collaborative Filtering (NeuralCF) (He et al. 2017), a graph convolution network model (LightGCN) (He et al. 2020), and a self-supervised learning model (SSL4Rec) (Yao et al. 2021).⁸ The main paper presents results using the BPR recommendation algorithm; the results for other techniques are presented throughout the Online Supplement due to page limits.

Simulation Assumptions. Simulation is a simplified version of the real-world environment. Many factors may affect real users' consumption choices and RS performance, but it is not feasible for an agent-based model to consider and quantify every possible factor. Here are some of the key assumptions that our simulation makes about the users and the environment: (1) each user consumes a given item at most once, i.e., no repeated consumption; (2) there are no monetary factors involved when users select items; (3) items do not have inventory restrictions and can be simultaneously consumed by multiple users (e.g., as is the case with most digital goods); (4) all recommendations are generated by a single RS; (5) all users have the same level of reliance on RS; (6) in the dynamic user preference setting, all users' preferences change homogeneously at the same time; and (7) users who are affected by the data privacy regulations are randomly chosen.

5 Experiment Results

5.1 Experiment 1: Impact of Data Privacy Practices with Stable User Preferences

Our main objective is to study how different data practices under privacy regulations impact the RS performance. The first experiment focuses on scenarios where the user population exhibits relatively stable preferences. Several traditional RS application domains, e.g., where taste-based goods (movies, books, music, etc.) having highly subjective preferences are recommended, can be considered as having more stable user

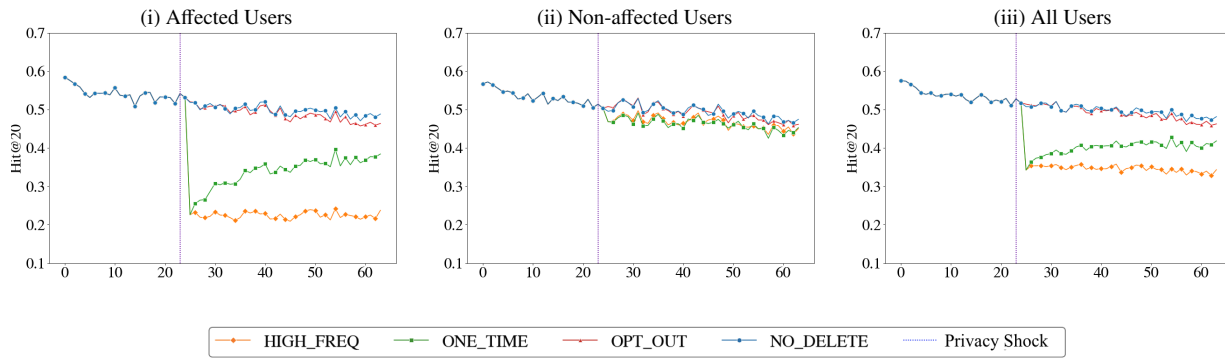
⁷Due to page limits, we present the information about these two additional datasets and the main results of our computational experiments on these datasets in Appendix E.

⁸We calibrate implementations of RS algorithms by benchmarking against reported results in their original papers. Benchmarking results are provided in Appendix F. Details of the hyper-parameter tuning process for these algorithms are provided in Appendix G. We use *random fallback* as the default fallback method, i.e., we recommend randomly selected items when the intended RS algorithm is unable (e.g., due to insufficient data) to produce recommendations for a given user.

preferences (Shen and Ball 2011). In our simulations, stable user preferences means that each user u 's actual preferences e_u remain the same throughout the entire period, i.e., $PrefList_u$ does not change.

Before the data privacy requests take effect at $T_{DataReq}$ (i.e., the 24th week in our simulations), there are no restrictions for the RS to collect users' data. Thus, we assume that the RS can take full advantage of all users' consumption data. After $T_{DataReq}$, we apply different data privacy requests $DataRequest_u$ on the user population, such as NO_DELETE, ONE_TIME, HIGH_FREQ, and OPT_OUT. The first set of results, presented in Figure 2, shows the RS performance over time for BPR in the presence of data privacy requests.⁹ All other simulation parameters are held constant (i.e., precisely controlled), including system initialization, the portion of affected users (50%), the degree of user reliance on RS (20%), etc., as discussed in Section 4. Furthermore, Figure 2 presents the RS performance results separately for the affected (column i) and non-affected (column ii) user sub-populations, as well as for the overall user population (column iii).

Figure 2: Impact of Data Privacy Practices with Stable User Preferences



Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%. The RS algorithm is BPR.

Because the simulation parameters are carefully controlled across different simulation variations, there is no difference in the RS performance before $T_{DataReq}$; the differences appear only after $T_{DataReq}$. Overall, the general RS performance patterns are highly consistent across RS algorithms (as shown in Appendix H). For the affected users, with no privacy requests in effect (i.e., NO_DELETE), the RS performance after $T_{DataReq}$ remains similar to what it was before, as shown in column (i) of Figure 2. The RS performance after the $T_{DataReq}$ can be summarized as: NO_DELETE $\approx$ OPT_OUT $>$ ONE_TIME $>$ HIGH_FREQ. That is, OPT_OUT does not influence the RS performance significantly over time, compared with NO_DELETE, because RS has already accumulated sufficient data before $T_{DataReq}$ to infer user preferences. Thus, when user preferences are stable, not having consumption data after $T_{DataReq}$ does not substantially harm the RS performance. In contrast, ONE_TIME and HIGH_FREQ lead to dramatic RS performance reductions right after $T_{DataReq}$, because affected users delete their historical data at $T_{DataReq}$, so the amount of available training data from the affected user population is minimal at that time. With ONE_TIME, the system can accumulate new consumption data after $T_{DataReq}$; thus, the RS performance of the affected users gradually improves. By the 60th week, the RS performance for affected users under ONE_TIME improves substantially,

⁹Due to page limits, the complete results for all RS algorithms (ALS, BPR, ItemKNN, NeuralCF, LightGCN, and SSL4Rec) are provided in Appendix H. The qualitative performance patterns remain highly consistent across different algorithms. Furthermore, we provide the computational cost analysis of different RS algorithms under different privacy practices in Appendix I.

although not to the same levels as with NO_DELETE. Under HIGH_FREQ, once the RS performance for the affected user population drops after $T_{DataReq}$, it does not improve afterward, as the available training data from this population remains minimal over time due to continuing periodic data deletions. Overall, different data privacy practices lead to differential impacts on the RS performance for affected users.

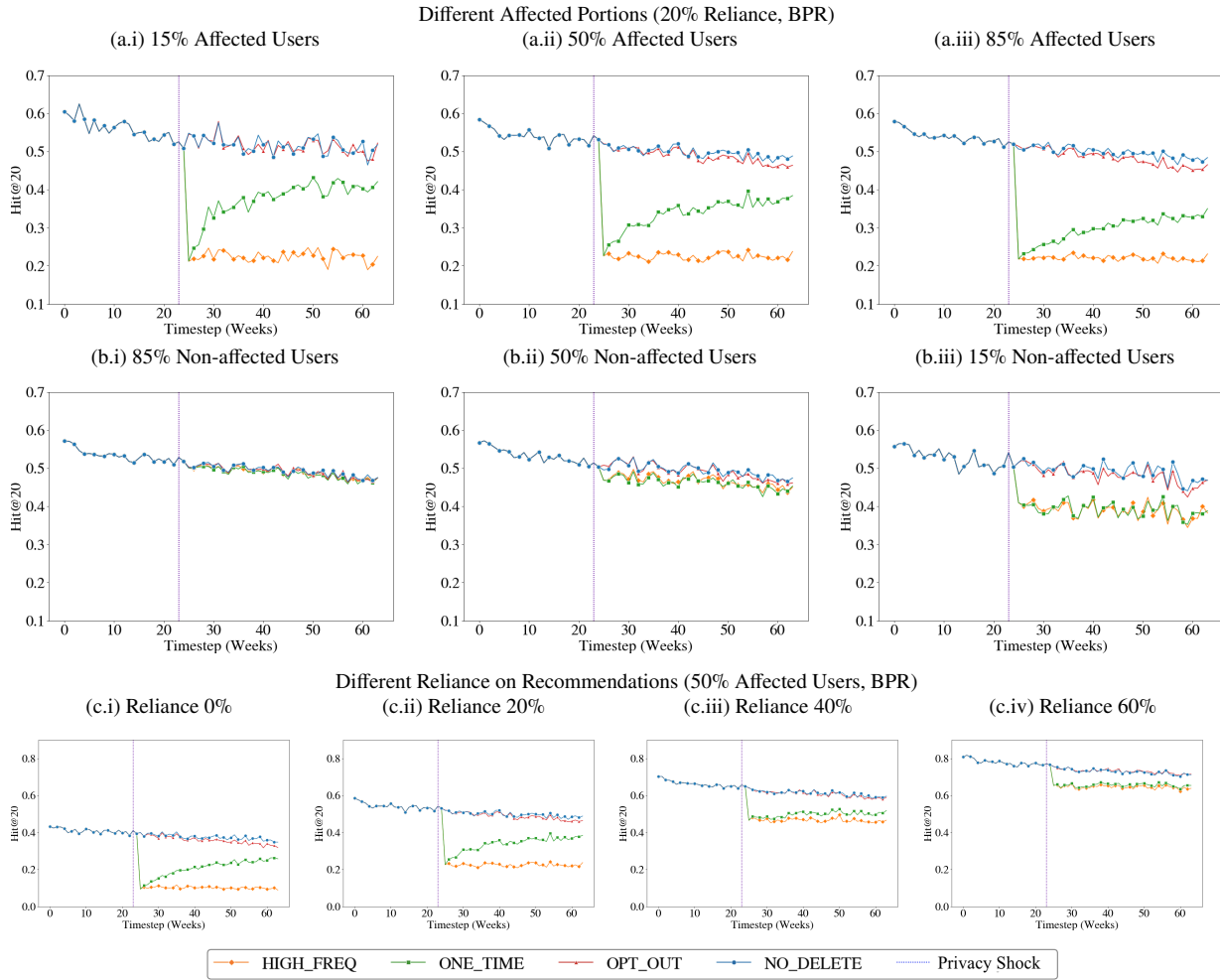
For the non-affected sub-population (i.e., users that do not request data practices), all historical and newly generated consumption data is available to RS. However, since RS integrates the data from the affected and non-affected users together for model training, losing data from affected users may affect the overall predictive ability of RS and, thus, may impact the RS performance for the non-affected users. Column (ii) of Figure 2 shows that there is almost no spillover effect of the data privacy practices from affected users to non-affected users across most RS algorithms. I.e., after $T_{DataReq}$, the RS performance for the non-affected users under different data privacy practices is similar to what it was before $T_{DataReq}$. This implies that RS still has sufficient data and inference power for the non-affected users, even when the other half of users request data removal in various ways. However, certain algorithms, i.e., LightGCN (as can be seen in Appendix H), can exhibit a substantial spillover effect, possibly because the underlying graph-learning method uses node embedding propagation on the user-item interaction graph, and the non-affected users' embedding learning is substantially harmed when the connected network becomes sparser after the data deletion of the affected users (Alon and Yahav 2020). Column (iii) of Figure 2 shows RS performance of the entire population, which is, as expected, the average performance of the affected and non-affected users.

To check the robustness of the results, we examine the impacts from two other factors, i.e., the proportion of the affected users and the extent of users' reliance on RS. We focus on the outcomes for the affected users, as they are impacted the most. First, we vary the portion of the affected users as 15%, 50%, and 85%, with the remainder being unaffected and fix users' reliance on RS at 20%. Second, we vary the degree of users' reliance on RS as 0%, 20%, 40%, and 60%, while keeping the portion of the affected users at 50%.

Figures 3(a.i-iii) show that the impact of data privacy practices is qualitatively similar across different portions of affected users. Only the RS performance trajectories with 15% affected portion are "noisier" because the total number of affected users is smaller, leading to a larger variance in performance. Also, under ONE_TIME, the RS performance recovers slower as the ratio of affected users increases, which is due to the loss of more historical data. For the non-affected users, the spillover effect from the affected users' data privacy practices changes non-linearly. When the portion of affected users is 15% (Figure 3(b.i)), there is almost no spillover effect, and with 50% (Figure 3(b.ii)) it is still negligible. Only when the portion of the affected users is large (e.g., 85%, as in Figure 3(b.iii)), there is a significant spillover effect, although the performance decrease for non-affected users is much smaller than that for affected users.

Regarding users' reliance on recommendations, Figures 3(c.i-iv) show that the impact of data privacy practices is qualitatively similar across different reliance levels. Quantitatively, the increase in the user's reliance on recommendations boosts the RS base performance (i.e., under NO_DELETE), because users are more inclined to consume the recommended items regardless of whether they reflect users' true preferences. Also, in addition to the increase in the base performance, performance loss under HIGH_FREQ and ONE_TIME practices gets smaller. This is because the performance loss from data privacy practices is typically due to having less information about user preferences. When information about user preferences is

Figure 3: Consistent Patterns across Affected Portion and Recommendation Reliance (Stable Preference)



less important for achieving good prediction performance (i.e., for populations with high reliance on RS), the negative impact of the data privacy practice also gets less substantial.

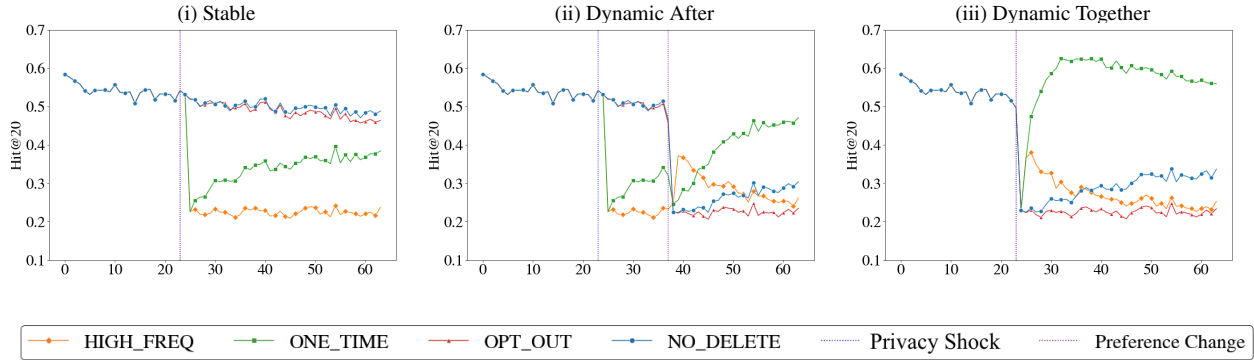
To summarize, Experiment 1 shows that, when user preferences are stable, the OPT_OUT practice creates a much smaller negative impact on the RS performance than the options of deleting historical data (e.g., ONE_TIME, HIGH_FREQ). The impacts are mainly on the affected users. And the impact of data privacy practices is a general, data-driven phenomenon, i.e., the qualitative impact patterns (such as relative performance impacts of different practices) are highly consistent across a diverse set of recommendation algorithms (as well as datasets, as shown in Appendices E.1 and E.2). Of course, the exact quantitative changes can vary across algorithms, as is generally the case with most machine learning tasks; exploring the algorithm-specific nuances of data privacy practices represents an interesting direction for future research.

5.2 Experiment 2: Impact of Data Privacy Practices with Dynamic User Preferences

Experiment 1 examined scenarios where user preferences remained stable over time. However, user preferences are not always static. In certain contexts, e.g., those involving functional, utility-based, or convenience goods, user preferences may fluctuate based on the users' current needs and circumstances. E-commerce platforms

represent one example setting where users could have variable, need-driven preferences (Swarnkar and Kalla 2021). Therefore, in Experiment 2, we explore how the data privacy practices impact the RS performance when user preferences can change.

Figure 4: Impact of Data Privacy Practices with Dynamic User Preferences (Affected Users)



Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%. The RS algorithm is BPR.

Extreme User Preference Change. In our initial investigation, we focus on an extreme and simplified scenario where all users' preferences undergo significant changes. We model this by replacing each user's preference profile with a preference profile from another, randomly selected user at some pre-specified time $T_{PrefChange}$. As discussed in Section 4, formally this is achieved by setting $\alpha = 1$ in Equation 5. To show whether the relative timings of the preference change and the data privacy request can influence the results, we explore two different settings: (i) $T_{PrefChange}$ is at Week 24, i.e., *together* with the data privacy request ($T_{DataReq} = T_{PrefChange}$), or (ii) $T_{PrefChange}$ is at Week 38, i.e., *after* the data privacy request ($T_{DataReq} < T_{PrefChange}$).¹⁰ In our experiments, we investigate the effects of a single, one-time preference change that occurs across the entire user population.

We first focus on the affected user population, and compare the impact of data privacy practices in the settings characterized by stable vs. extremely dynamic user preferences (i.e., the degree of preference change $\alpha = 1$). The results, provided in Figure 4, highlight that different data privacy practices have substantially different impacts.¹¹ For reference, column (i) of Figure 4 shows the impact of data privacy practices in the context of stable user preferences, which is the same as the first column of Figure 2. Column (ii) of Figure 4 represents a scenario where user preferences change *after* the data privacy request. When RS operates normally and there are no data privacy requests (i.e., NO_DELETE), after users' preferences change at $T_{PrefChange}$, the RS performance decreases immediately and then increases slightly afterwards. This validates that the simulation of preference change is effective – the RS performance drops because the historical consumption data before preference change is no longer helpful for inferring users' (new) preferences. By comparing columns (ii) and (iii), we observe that the effects of preference changes are qualitatively consistent regardless of the preference change timing (i.e., whether $T_{PrefChange}$ is after or together with $T_{DataReq}$).

First, in stable preference settings, the RS performance under OPT_OUT is similar to that of NO_DELETE,

¹⁰The preference change can also happen before the data privacy request. In that case, the result would be qualitatively similar to the stable preference scenario, since the RS performance starts to differ only after data privacy request.

¹¹Due to page limits, the complete results for all RS algorithms (ALS, BPR, ItemKNN, NeuralCF, LightGCN, and SSL4Rec) are provided in Appendix K. The qualitative performance patterns remain largely consistent across different RS algorithms.

as discussed earlier. In contrast, when user preferences change significantly at $T_{PrefChange}$, the newly generated consumption activities are much more valuable for inferring new preferences; thus, not collecting the newest activities (i.e., with OPT_OUT) significantly harms the RS performance, i.e., it cannot gradually improve over time, as compared to NO_DELETE.

Second, looking at the scenario where $T_{PrefChange} = T_{DataReq}$ (column (iii)), the RS performance under deletion options (ONE_TIME and HIGH_FREQ) substantially improves immediately after $T_{PrefChange}$. The main reason is that the historical (pre- $T_{DataReq}$) consumption data becomes much less relevant with preference changes, and dropping it can substantially improve the RS performance. Moreover, under ONE_TIME, RS can accumulate users' newly generated consumption data for model updating, so the RS performance keeps growing. Thus, the longitudinal performance under ONE_TIME can be substantially higher than under NO_DELETE because of the removal of historical (and less relevant, in this case) data.

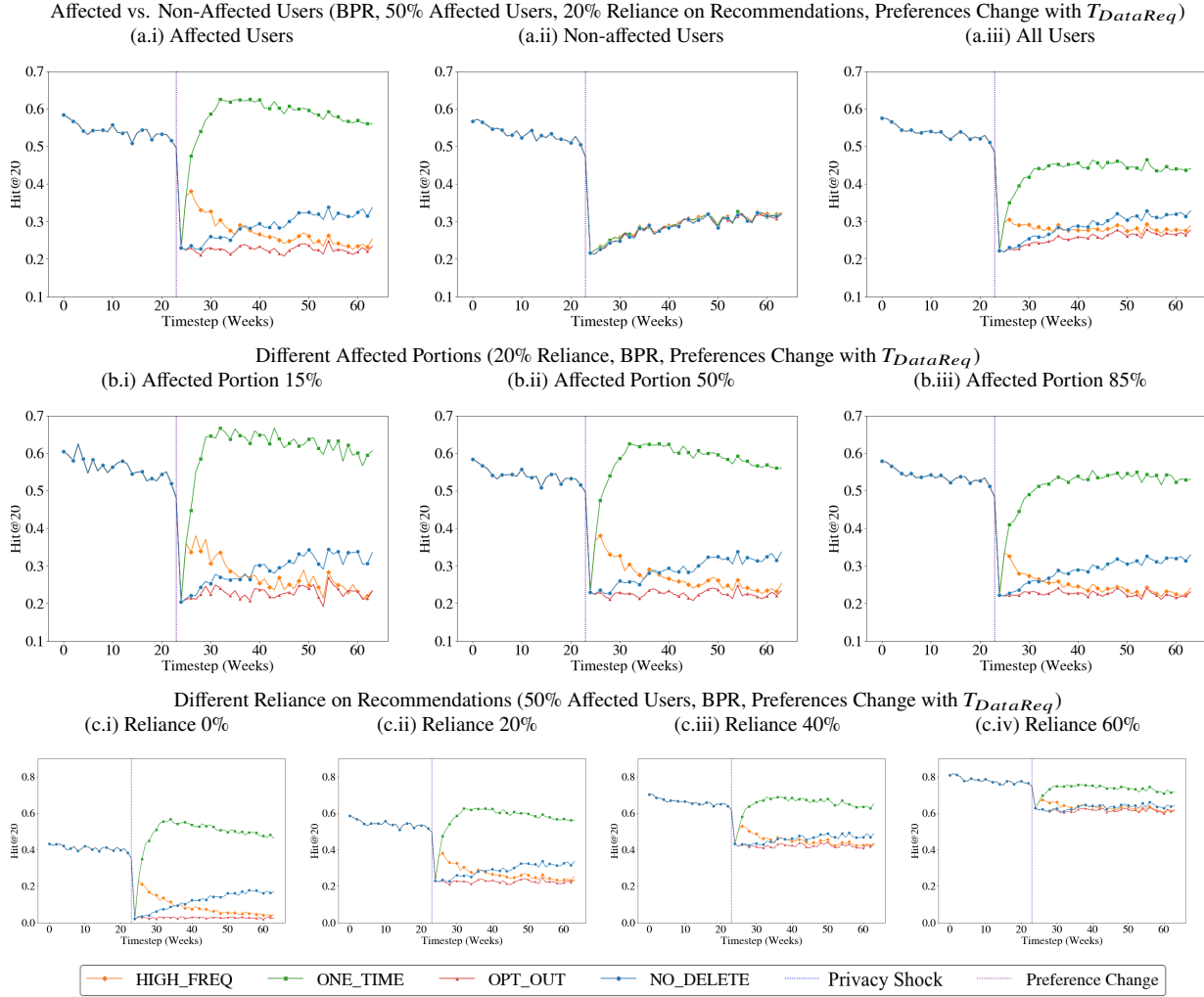
Third, comparing columns (ii) and (iii) of Figure 4, i.e., comparing settings where $T_{PrefChange}$ is temporally separate from vs. together with $T_{DataReq}$, we can observe some additional nuances of deletion-based data privacy practices. Specifically, when $T_{DataReq} < T_{PrefChange}$ (i.e., column (ii)), the RS performance under HIGH_FREQ exhibits a very substantial jump immediately after $T_{PrefChange}$ for several time intervals, outperforming even ONE_TIME during that time. The explanation is twofold: (1) ONE_TIME has already been able to accumulate some historical data after $T_{DataReq}$ that has now (i.e., after $T_{PrefChange}$) become less relevant, whereas HIGH_FREQ always uses only the most recent data; (2) it is important to note that, after $T_{PrefChange}$, the first few most relevant items tend to be more “distinctive” and relatively easier for RS to predict even with smaller amounts of user data, as we have under HIGH_FREQ. As the most relevant items get consumed by users, the relevance of the unconsumed item candidates gets lower and less distinctive, and it becomes more difficult for the RS to make accurate top-N item predictions with more homogeneous items in terms of relevance, especially with small amounts of user data.¹² This is the reason for the gradual decrease of HIGH_FREQ performance after its initial post- $T_{PrefChange}$ peak.

Similar to Experiment 1, in Experiment 2 we focus mainly on the results for the affected user population, because applying different data privacy practices does not result in significant RS performance differences for non-affected users, as can be observed in Figure 5(a.ii). In addition, we observe that the qualitative performance patterns in dynamic preference settings (that are quite different from performance patterns in stable preference settings) remain largely consistent for different portions of affected users (as shown in Figures 5(b.i-iii)) and different degrees of users' reliance on RS (as shown in Figures 5(c.i-iv)).

Different Degrees of User Preference Change. To show the impact of data privacy practices for different degrees of user preference change, we vary α in Equation (5) as 0 (i.e., stable preferences), 0.2, 0.4, 0.6, 0.8, and 1.0 (i.e., the extreme preference change). As we have observed qualitatively consistent RS performance patterns by varying $T_{PrefChange}$, recommendation algorithms, different portions of affected users, and reliance on recommendations, we therefore fix these factors in this variation of simulation experiments. The results are presented in Figure 6. Under NO_DELETE, as the degree of preference change (i.e., α) increases, the RS performance after $T_{PrefChange}$ becomes worse, since the historical data is increasingly less helpful for inferring users' new preferences. Also, the RS performance under OPT_OUT is worse than that under

¹²Please see Appendix J for a more in-depth explanation.

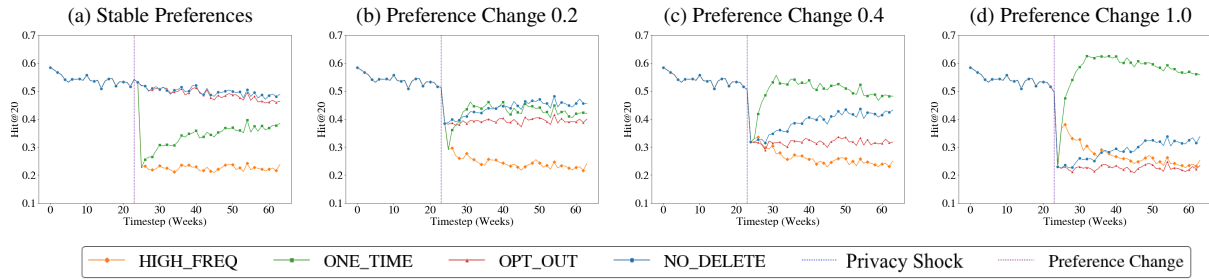
Figure 5: Consistent Patterns across Affected Portion and Recommendation Reliance (Dynamic Preference)



NO_DELETE as the degree of preference change increases. This is a combined effect of both (a) inability to use new data generated after the data privacy request under OPT_OUT and (b) the old data retained from before the data privacy request not reflecting the new user preferences, as the degree of preference change increases. On the other hand, the RS performance under ONE_TIME or HIGH_FREQ practices improves as the degree of preference change increases, consistent with our earlier observations for the case of extreme preference changes. Moreover, we observe that the RS performance under ONE_TIME can be especially advantageous in contexts where user preferences can change – the larger the degree of user preference change, the less relevant the historical consumption data, and the more useful the deletion of historical data becomes. This exploration further validates the effectiveness of the proposed approach for simulating scenarios where user preferences change and provides additional important insights about the impact of various data privacy practices. Also, the results in Figure 6 demonstrate the gradual and consistent change of impacts from different data privacy practices as the user preference change increases.¹³

¹³Appendix L shows more extensive results of this robustness check, with additional degrees of preference change included (i.e., 0.6 and 0.8). The gradual changes in RS performance are consistently observed across different degrees of preference change.

Figure 6: Impact of Varying Levels of User Preference Change



The results are on affected users and the simulation is under a setting where 50% of users are affected, the RS algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Overall, for recommendation platforms where user preferences are less dynamic, opting out of data collection impacts the RS performance substantially less than the deletion-based alternatives. However, in scenarios where user preferences are more volatile, deletion-based options (especially deletion with low frequency, e.g., ONE_TIME) are much more advantageous than opt-out alternatives.

5.3 Experiment 3: Impact of Data Privacy Practices with Incremental Learning

Experiments 1 and 2 explore the impact of data privacy practices on the RS performance and focus on the situations where consumers are able to submit data privacy requests pertaining to their personal data collection by online platforms. However, as mentioned in Section 3, there is no uniform stance among current privacy regulations on whether *inferences* (e.g., embeddings, latent representations) learned about users by machine learning models should be explicitly designated as a category of personal data. Therefore, for the situations where learned inferences may not fall under personal data protections, Experiment 3 investigates the potential implications of the *incremental learning* approach that can utilize these learned inferences in order to improve the RS performance under data privacy practices.

More specifically, the incremental learning approach has access to the previously trained model, which includes model parameters as well as learned inferences (e.g., latent representations) about users and items. A key beneficial characteristic of this approach is the ability to incorporate new data *as it becomes available* and update the pre-existing model gradually over time, eliminating the need to store historical data; this is in contrast to utilizing the entire currently available data for model building all at once, i.e., each time rebuilding the model from scratch. However, incremental learning is not always able to match the prediction performance of traditional, non-incremental learning algorithms because of *forgetting*; i.e., when an incremental learning algorithm updates the model using new incoming data, the new updates can have a disproportionate effect in that they can override knowledge acquired from previous data (Liu et al. 2020, Pereira 2022).¹⁴ In summary, it is not apparent whether incremental learning would be advantageous at addressing the challenges brought by data regulations, and we explore this in Experiment 3.

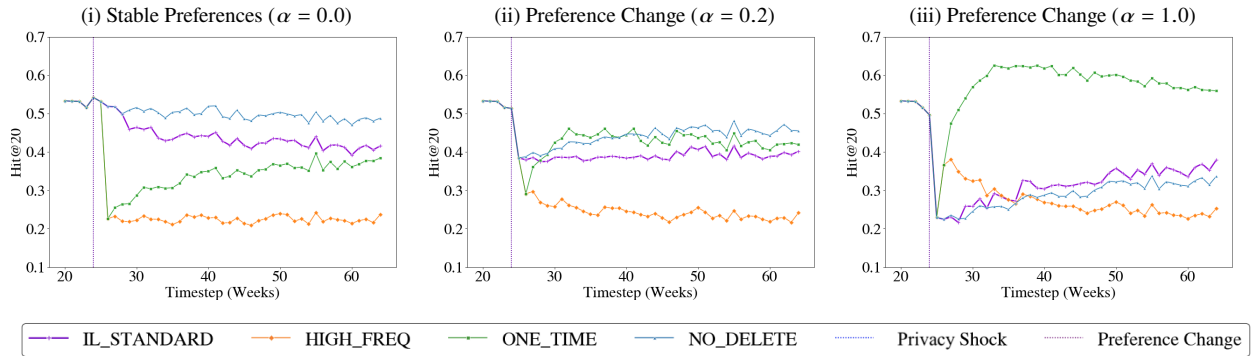
In Experiment 3, we use BPR as a representative example of a recommendation algorithm that can be used in an incremental learning manner. In our simulations, we assume that platforms would use non-

¹⁴Thus, incremental learning algorithms have to be fine-tuned well for a specific application domain in order to curtail forgetting-related performance limitations as much as possible.

incremental-learning RS algorithms before the data privacy request $T_{DataReq}$, as access to entire historical data is beneficial from the RS performance perspective. After $T_{DataReq}$, RS keeps the latest model, including user and item latent representations (profiles), and incrementally updates them going forward.

Because incremental learning does not rely on the entire historical data for model updating (i.e., all new data is incorporated periodically, as it arrives, and can then be deleted), its performance under the two deletion practices (ONE_TIME and HIGH_FREQ) is equivalent to that of the standard incremental learning approach. We call this setting “standard incremental learning” (IL_STANDARD), meaning all users’ consumption data that arrived after the previous model update is used for incremental learning updates. Under OPT_OUT, the affected users’ new consumption data is not available; thus, only non-affected users’ new consumption data can be used for incremental updates. We call this setting “incremental learning with opt out” (IL_OPT_OUT). To better understand the impact of the incremental learning capabilities, we use BPR to separately compare each of these two settings to their corresponding non-incremental counterparts. When presenting the results, we continue to use ONE_TIME, HIGH_FREQ, OPT_OUT, and NO_DELETE to denote the performance of non-incremental-learning BPR under these data privacy practices. Moreover, to show the important role of the changes in user preferences, we vary the degree of the user preference change (i.e., α in Equation (5)) as 0 (stable preferences), 0.2, and 1.0 (the extreme preference change).¹⁵

Figure 7: Incremental and Non-incremental Learning under Deletion-Based Practices (Affected Users)



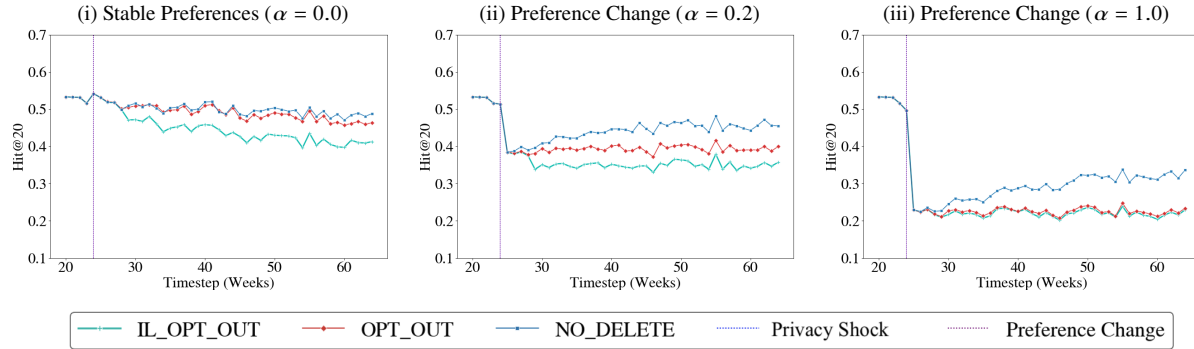
The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

We first compare incremental and non-incremental learning under deletion-based data privacy practices, focusing on affected users, as shown in Figure 7. Usefulness of incremental learning varies based on the degree of user preference change. When user preferences are stable, IL_STANDARD performs worse than NO_DELETE, due to the negative effect of forgetting (Figure 7(i)). However, IL_STANDARD outperforms the system under ONE_TIME and HIGH_FREQ (Figure 7(i)), indicating that incremental learning can help preserve the RS performance under deletion-based practices with stable user preferences. When user preferences can change significantly (Figure 7(iii)), longitudinal performance of IL_STANDARD continues to outpace that of HIGH_FREQ. Moreover, IL_STANDARD (slightly) outperforms NO_DELETE as well, because forgetting the old user preferences can be beneficial for inferring the new user preferences after

¹⁵We explored the variations where $\alpha = 0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0$ – relative performance positions of different data privacy practices change continuously. Also, the qualitative patterns when $\alpha \geq 0.4$ are highly similar. Therefore, we chose the two boundary scenarios ($\alpha = 0$ and $\alpha = 1.0$) and one intermediate scenario ($\alpha = 0.2$) to illustrate the gradual change.

$T_{PrefChange}$. At the same time, the performance of IL_STANDARD and NO_DELETE is substantially inferior to that of ONE_TIME after a significant preference change (Figure 7(iii)), because taking into account the pre- $T_{PrefChange}$ user data (that became much less relevant) – either directly (with NO_DELETE) or via latent representations (with IL_STANDARD) – adds noise in model training and harms the RS performance.

Figure 8: Incremental and Non-incremental Learning under Opting Out Practice, Hit@20 (Affected Users)



The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

We also compare incremental vs. non-incremental learning when users request opting out of data collection (i.e., IL_OPT_OUT vs. OPT_OUT), focusing on affected users, as shown in Figure 8. Overall, incremental learning does not provide substantial performance benefits under opt-out-based data privacy practices, regardless of the user preference dynamics. As discussed earlier, when user preferences are stable, OPT_OUT does not substantially affect the RS performance (compared to NO_DELETE). Meanwhile, IL_OPT_OUT can harm the RS performance due to the forgetting effect, as illustrated by Figure 8(i). When user preferences can change dramatically, the RS performance under IL_OPT_OUT and OPT_OUT is very similar for the affected users, as shown in Figure 8(iii).¹⁶

Overall, incremental learning represents a viable strategy for reducing the performance impacts of deletion-based data privacy practices – it provides consistent improvements over HIGH_FREQ, but is advisable as an alternative to ONE_TIME only when user preferences are relatively stable. For opt-out-based practices, incremental learning does not provide RS performance benefits as compared to the non-incremental-learning counterparts.

5.4 Robustness Checks and Additional Analyses

To assess the robustness and generalizability of our simulation framework, we conducted several supplementary analyses. As mentioned earlier, our main results are robust across a number of RS algorithms (Appendices H and K), datasets (Appendix E), accuracy metrics (Appendix D), and key simulation parameters (Appendix C). Furthermore, Appendix O presents the main results from our experiments using confidence-band-based visualizations across 10 simulation runs, confirming that observed performance differences among data privacy practices are consistent and substantial. As a further illustration of performance differences between data privacy practices in our experiments, Appendix M presents statistical tests of post-privacy-shock performance

¹⁶In addition to the main results for affected users provided in Figures 7 and 8, for completeness we also include the corresponding results for non-affected users and for the entire user population in Appendix N.

differences. We also perform additional robustness checks on the external validity of the simulation in Appendix Z, where we benchmark the simulated consumption behavior against seeding (real-world) data on user activity and item popularity; this provides further support regarding the realism, relevance, and robustness of the modeling approach.

In addition to modeling a sudden and homogeneous shift in user preferences, we also explore gradual and heterogeneous changes. Specifically, Appendix P analyzes a scenario where the user preference change occurs gradually over time. Appendix Q investigates a scenario in which the degree of preference change is individualized for each user, based on their consumption history. In addition, Appendix R further examines heterogeneous dynamic user preferences, where the degree of preference change is item-category-specific, i.e., it depends on the item that the user consumes. Also, Appendix S explores a scenario with staggered timing of user preference changes, i.e., representing non-simultaneous privacy “shocks”. We also relax the simulation assumption of a homogeneous and static user reliance on recommendations. In particular, heterogeneous (and calibrated) user reliance on RS is explored in Appendix T by estimating and seeding reliance values derived from the click-stream logs from real-world data. Also, Appendix U introduces a model for heterogeneous and dynamic (i.e., time-varying) user reliance on RS, where it is adjusted in an individual manner based on the observed RS performance. To build additional realistic considerations into the simulation, Appendix V incorporates new item arrivals, utilizing an arrival schedule that reflects new item dynamics observed in real-world data. Across all these robustness checks and extensions, we find that the patterns of long-term RS performance and the relative impact of data privacy practices remain highly consistent with our original findings, even under more granular modeling of user and item dynamics.

In addition to aforementioned robustness checks, we also evaluate broader item consumption outcomes influenced by data privacy practices. Appendix W analyzes item consumption diversity and shows that deletion-based practices substantially reduce diversity, while OPT_OUT preserves it, and these patterns persist across stable vs. dynamic user preference settings. Appendix X examines long-tail item consumption and reveals that, in extreme low-data situations, the long-tail consumption outcomes vary primarily based on what “fallback” recommendation strategies (e.g., random vs. popular item recommendation) are being used by the RS algorithms. These insights are further validated in Appendix Y, where additional *hybrid* fallback strategies are considered and their impact on key metrics of system performance and consumption dynamics is assessed. Finally, we observe that data privacy practices have minimal spillover effects on non-affected users in terms of consumption diversity and longtailness.

6 Parameterizable Privacy Practices based on Revelation Protection

Although existing data privacy practices (as discussed in Table 1 and indicated in various data privacy regulations) provide several options for platforms and consumers, they typically represent relatively “blunt”, all-or-nothing instruments of controlling individual users’ data collection that have significant effect on the RS performance. Thorough understanding of personalization performance implications of these existing, canonical data privacy practices (as discussed in the preceding experiments) provides rich opportunities for designing more flexible, parameterizable data privacy practices that allow consumers and platforms to

navigate the personalization-privacy tradeoff in a more nuanced manner.

One natural design direction towards flexibility and parameterizability would be to consider *partial* data privacy practices, e.g., only deleting (or opting out of) *portions* of individual user data. One key challenge is how to design a meaningful parameterization of the data space in personalization-privacy-aware manner, so that the deletion (or opting out) of a data portion can increase privacy benefits as much as possible while reducing personalization benefits as little as possible. In this section, we propose one example approach to parameterizing data privacy practices based on the notion of *revelation protection* and demonstrate its effectiveness using a set of computational experiments. The key intuition behind the proposed data privacy practices is that some portions of an individual user’s data can more strongly differentiate this user from others in the population and, hence, are potentially more revealing about the user’s uniqueness and identity.

The example approach that we propose and investigate in this paper uses *item (un)popularity* as an informative proxy for information revelation. Intuitively, consumption of *niche* (i.e., less popular) items is indicative of more unique preferences; thus, such information (unlike information about the popular item consumption) would be much more helpful when distinguishing a given user from the others. Consequently, we propose enhanced versions of the canonical privacy practices (i.e., based on data deletion or opting out) that can target parameterizable portions of data pertaining to less popular items.¹⁷ This approach provides platforms and users with finer levels of control over the quantity of data withheld and is intended to navigate the personalization-privacy tradeoff more effectively, which is strongly supported by the empirical results provided in this section. In other words, substantial privacy benefits for a given consumer can already be achieved by removing a smaller amount of data on niche item consumptions (and not having to remove the entire user’s consumption history), while retaining substantial personalization benefits from the remaining consumption data (leveraging *collaborative filtering* capabilities of modern, state-of-the-art RS algorithms).

Formally, we define I_{niche}^c as the portion $c \in [0, 1]$ of items with the lowest overall consumption counts, and we apply data privacy practices only to I_{niche}^c . I.e., for RP_ONE_TIME and RP_HIGH_FREQ, only consumptions of items from I_{niche}^c are deleted; similarly, for RP_OPT_OUT, we only opt out of collecting data on items I_{niche}^c . “Niche” threshold c specifies how much data is under the control of the enhanced data privacy practices: $c = 0$ represents NO_DELETE, $c = 1$ represents the full canonical versions of the data practices, and the intermediate c values provide a more granular control over the individual users’ data.

To quantify the extent to which the data collected by RS *reveals* the uniqueness of a given user’s preferences, we define *Mean Pairwise Revelation* (MPR) for a focal user u at a given timestep t as follows:

$$\text{MPR}_t(u) = \frac{1}{|U| - 1} \sum_{v \in U \setminus \{u\}} \text{PR}_t(u|v) = \frac{1}{|U| - 1} \sum_{v \in U \setminus \{u\}} \sum_{i \in D_{u|v}^t} \text{InvPop}_t(i). \quad (6)$$

Here $\text{PR}_t(u|v)$ is *Pairwise Revelation* of the focal user u with respect to another user v at time t . More formally, $\text{PR}_t(u|v) = \sum_{i \in D_{u|v}^t} \text{InvPop}_t(i)$, where $D_{u|v}^t = \text{TrainData}_u^t \setminus \text{TrainData}_v^t$ denotes the set of items consumed by user u but not by v at time t (i.e., items that differentiate u in this pair of users), and $\text{InvPop}_t(i)$ indicates the inverse popularity (i.e., inverse aggregate consumption count) of item i at time t . Thus, at its

¹⁷We will denote these parameterizable versions with prefix “RP_”, for “Revelation Protection”. For instance, RP_ONE_TIME denotes a parameterizable version of the canonical ONE_TIME data privacy practice, and so on.

core, $MPR_t(u)$ reflects the uniqueness (or revelation) of user u 's available item consumption data as compared with other users' consumption data. Finally, as a global revelation measure for the entire user population, we calculate $MPR_t = \frac{1}{|U|} \sum_{u \in U} MPR_t(u)$. Higher MPR values indicate that, on average, based on the item consumption data available to RS, users are more distinguishable from one another, while lower MPR values reflect more privacy-friendly settings.

Experiment 4. Figures 9(a.i-iv) show the longitudinal impact on the global MPR metric under different privacy practices with stable user preferences¹⁸, for the increasing values of c which reflect the users' ability to control increasingly larger portions of personal data. Unsurprisingly, without any data privacy practices (i.e., under NO_DELETE), MPR remains highest in all settings over time, as larger (and increasing) consumption histories make it easier to differentiate users from one another. Compared with NO_DELETE, MPR of RP_HIGH_FREQ and RP_ONE_TIME decreases significantly after the privacy shock, indicating that it is much more difficult to differentiate between users with limited data. Moreover, MPR under RP_HIGH_FREQ stays low afterwards, as users keep deleting personal data periodically, while MPR under RP_ONE_TIME increases gradually because of the continuous accumulation of personal data after the initial deletion. In contrast, MPR under RP_OPT_OUT gradually decreases. Users stop sharing their new consumption data under RP_OPT_OUT, but the popularity of their already consumed items can only increase over time. As these previously consumed items become more common, the decreasing MPR under RP_OPT_OUT appropriately reflects decreasing uniqueness of user consumptions over time. Furthermore, the relative MPR performance patterns for different data privacy practices are highly consistent across varying c values; and, in absolute terms, as expected, the more data users choose to control using the proposed parameterizable data privacy practices (i.e., larger c), the less distinguishable users are from one another (i.e., lower MPR). Finally, analogous patterns also hold for the dynamic preference scenarios, as shown in Figures 9(b.i-iv). Therefore, MPR is an appropriate and robust measure for the preference revelation.

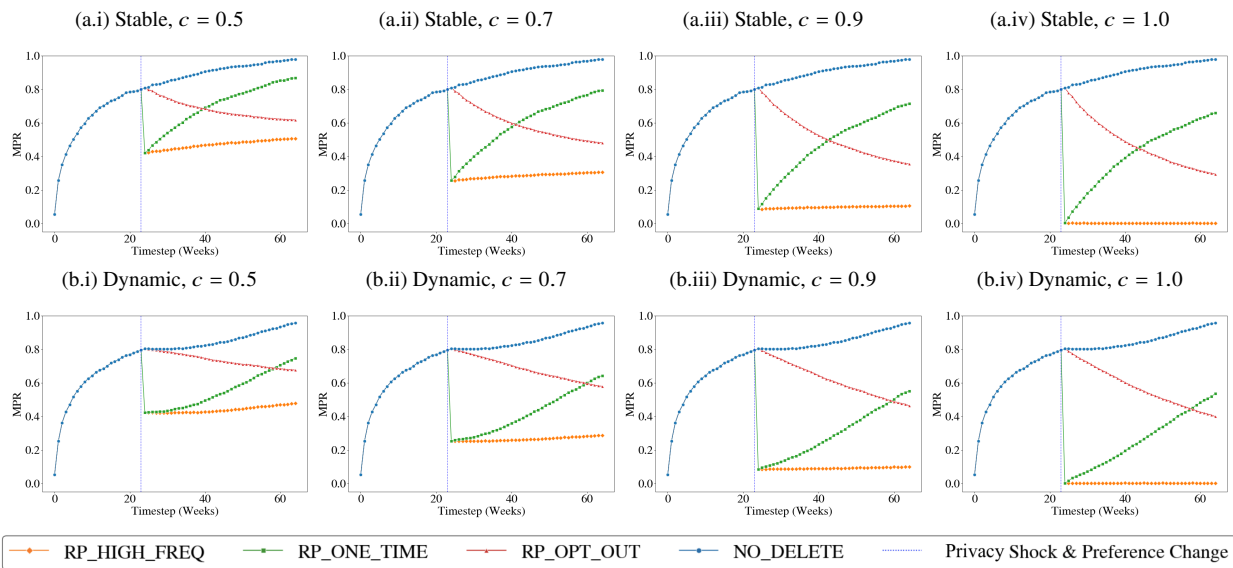
Importantly, we explore the proposed parameterizable data privacy practices in terms of their ability to navigate effectively the personalization-privacy tradeoff. In other words, could threshold c value be tuned in a way that provides substantial revelation protection, while at the same time retaining significant RS performance benefits? We vary c from 0.5 to 1 by 0.05 and evaluate the average MPR and $Hit@20$ across all timestamps after the privacy shock. Figures 10(a.i-iii) illustrate the tradeoff under the stable user preference scenarios. Under data deletion practices (i.e., RP_HIGH_FREQ and RP_ONE_TIME), a modest reduction in c from 1.0 to 0.9 (i.e., using only user consumption data from 10% most popular items) leads to a very substantial, near-optimal increase in $Hit@20$ (i.e., comparable to NO_DELETE performance) with only a small increase in MPR. In contrast, under RP_OPT_OUT, $Hit@20$ shows minimal sensitivity to changes in c , suggesting that more granular data control does not significantly enhance RS performance.

Figures 10(b.i-iii) present the tradeoff under the dynamic user preference setting. The patterns for RP_OPT_OUT are similar to patterns observed in the stable preference setting: a small reduction in c (e.g., from 1.0 to 0.9) can yield significant improvement in $Hit@20$ with only a small increase in MPR. However, the patterns for RP_ONE_TIME and RP_HIGH_FREQ are quite different. For example, in the

¹⁸Appendix AA shows the same statistics for dynamic user preferences. Patterns are highly consistent and readers are kindly referred to ther for details.

case of RP_ONE_TIME, since removing historical data improves the RS accuracy when user preferences are dynamic (as discussed earlier in the paper), the canonical configuration that deletes all historical data (i.e., $c = 1.0$) achieves both the highest *Hit@20* and the lowest MPR, thus, dominating the partial deletion configurations (i.e., when $c < 1.0$) on both dimensions. In the case of RP_HIGH_FREQ, we see an interesting non-monotonic relationship between c and RS performance, where focusing on consumption of only the 5% most popular items (i.e., $c = 0.95$) represents the optimal personalization performance, while retaining nearly all revelation protection benefits. The reason for this relationship is that, decreasing c from 1.0 to 0.95 allows RS to retain more data and escape the extreme low-data setting of standard HIGH_FREQ; however, decreasing c further raises the likelihood of including more “outdated” personal data (from before the preference change), which significantly impairs the RS algorithm’s ability to learn users’ true preferences.

Figure 9: Longitudinal MPR under Different Data Privacy Practices and Niche Thresholds c (Affected Users)



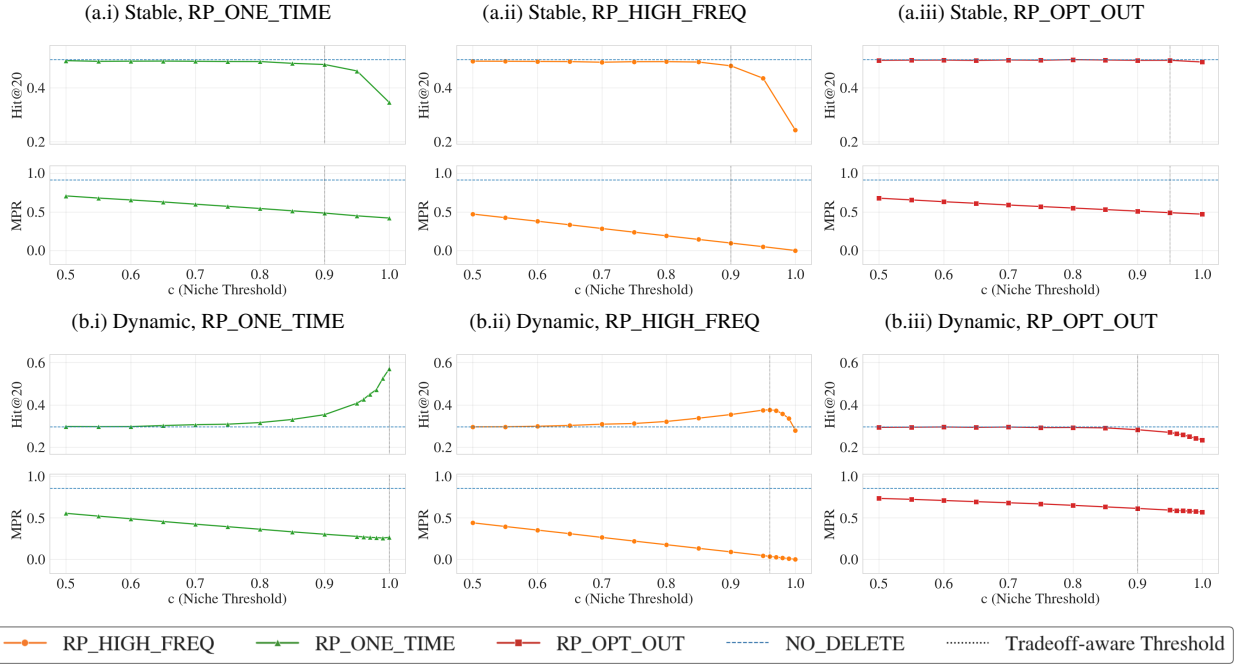
Across all figures, 50% of users are affected, user reliance on RS is 20%, and the algorithm used is BPR. Setting $c = 1.0$ represents canonical, fully enacted (i.e., non-niche-based) data privacy practices. Static user preferences correspond to $\alpha = 0$; dynamic preferences correspond to $\alpha = 1.0$.

In summary, the proposed parameterizable design of data privacy practices can help platforms and consumers achieve more nuanced, optimized benefits in both increasing personalization performance and reducing personal data revelation/unique-ness.

7 Conclusions

Overview of Findings. Our study systematically evaluates the performance implications of various data privacy practices in RS, with the key findings summarized in Table 2. These main findings together comprehensively document how user consumption choices (influenced by preference dynamism, reliance on RS, etc.) along with their privacy actions affect the RS training data (e.g., data size/availability and data relevance), through which their impact translates into RS performance. In addition, Table 3 summarizes extensive robustness checks and supplementary analyses that extend and validate these findings across multiple algorithms, datasets, evaluation metrics, and characteristics of user and item populations. The consistency of results across these dimensions underscores the generalizability and robustness of our findings.

Figure 10: Dynamics of MPR and Hit@20 under Varying Niche Thresholds (Affected Users)



Each dot in the plot represents the average outcome from 10 simulations under the setting where 50% of users are affected, user reliance on RS is 20%, and the algorithm used is BPR. Static user preferences correspond to $\alpha = 0$; dynamic preferences correspond to $\alpha = 1.0$.

Table 2: Main Findings: Impact of Data Privacy Practices on RS Performance

Data Privacy Practices	Under Stable Consumer Preferences	Under Dynamic Consumer Preferences
Canonical Data Privacy Practices: ONE_TIME delete HIGH_FREQ delete OPT_OUT	Deletion-based practices substantially reduce RS performance, with HIGH_FREQ having more severe impact than ONE_TIME. OPT_OUT is more advantageous in this scenario and has only negligible effect on RS performance. [Section 5.1 and Appendix H]	OPT_OUT severely impairs RS performance. Deletion-based practices are more advantageous in this scenario; in fact, ONE_TIME is a dominant approach, as it even (substantially) enhances RS performance, as compared to NO_DELETE. [Section 5.2 and Appendix K]
Incremental Learning Approaches: IL_STANDARD delete IL_OPT_OUT	Incremental learning (IL_STANDARD) reduces RS performance loss under deletion-based practices, outperforming both ONE_TIME and HIGH_FREQ. IL_OPT_OUT is highly ineffective, being inferior to the canonical OPT_OUT. [Section 5.3 and Appendix N]	IL_STANDARD can mitigate RS performance loss of HIGH_FREQ, but it is ineffective (or even harmful) as an alternative to ONE_TIME. RS performance differences between IL_OPT_OUT and OPT_OUT are negligible. [Section 5.3 and Appendix N]
Proposed Parameterizable Approaches based on Revelation Protection: RP_ONE_TIME delete RP_HIGH_FREQ delete RP_OPT_OUT	By tuning the data inclusion criterion (i.e., niche threshold c), RS can achieve near-best performance (i.e., comparable to NO_DELETE) for all data privacy practices, while retaining the vast majority of privacy-related benefits of these practices (i.e., reduction of personal data uniqueness/revelation). [Section 6]	By tuning the data inclusion criterion (i.e., niche threshold c), RS can achieve advantageous performance for all data privacy practices in this setting as well. In particular, deletion-based practices (RP_ONE_TIME and RP_HIGH_FREQ) are even able to outperform NO_DELETE. [Section 6]

Practical Implications. This study provides several practical implications for platform designers. First, in-depth analysis and evaluation of how various data privacy practices can affect RS performance is of direct relevance to the system designers. Based on the findings in this study, companies can strategically design their privacy choice architectures in ways that not only satisfy data protection regulations, but also provide

Table 3: Additional Findings (Robustness Checks and Supplementary Analyses)

Computational Settings	
Different RS algorithms	Main findings are consistent across different RS algorithms, including: ALS, BPR, ItemKNN, LightGCN, NeuralCF, and SSL4Rec. [Appendices H and K]
Different datasets	Main findings are consistent across different datasets, including the Multichannel Retailer, RetailRocket, and Amazon Book datasets. [Appendix E]
Different fallback methods	The choice of fallback methods (e.g., random, popularity-based, hybrid recommendations) affect mainly RS performance with HIGH_FREQ. Popularity-based fallback improves predictive performance at the cost of diversity and long-tail coverage. [Appendix Y]
Recommendation Quality Metrics	
Predictive accuracy metrics	Main findings are consistent across different accuracy metrics, including Hit@20, MRR@20, and NDCG@20. [Appendix D]
Item discoverability metrics: consumption diversity and longtailness	Deletion-based data privacy practices substantially reduce consumption diversity, while the opt-out practice does not. Biggest changes to long-tail item consumption happen under deletion-based practices in low-data situations (i.e., right after privacy shock) and are driven largely by the “fallback” strategies used by RS algorithms. [Appendices W and X]
User Population	
Non-affected users	The effect of data privacy practices on non-affected users (i.e., on users that choose not to make data privacy requests) is minimal in a vast majority of settings. That is, the overall effects are concentrated on the affected users. [Sections 5.1 and 5.2]
Portion of affected users	Main findings are consistent in user populations with different portions of affected users (i.e., 15%, 50%, and 85%). [Sections 5.1 and 5.2]
Reliance on recommendations	Main findings are consistent in user populations with different levels of homogeneous reliance on recommendations (i.e., 0%, 20%, 40%, and 60%) [Sections 5.1 and 5.2], with heterogeneous and calibrated user reliance on recommendations [Appendix T], and with heterogeneous and dynamic (time-varying) user reliance [Appendix U].
Recommendation list exploration	Main findings are consistent in user populations with different degrees of recommendation list exploration (i.e., β is 1.1, 1.3, and 1.7). [Appendix C]
Users' preference change	The gradual and consistent change of impacts from different data privacy practices on RS performance is observed, when varying the user population's homogeneous degree of preference change from stable to extreme ($\alpha = 0, 0.2, 0.4, 0.6, 0.8$ and 1.0) [Section 5.2 and Appendix L]. Main findings in the the dynamic preference setting remain consistent when modeling user preference change as gradual over time (instead of instantaneous) [Appendix P] and as staggered over time (instead of simultaneous) [Appendix S]. Main findings also remain consistent across under various models of heterogeneous user preference change: consumption-dependent [Appendix Q] and item-category-dependent [Appendix R].
Item Population	
New item arrival	Main findings are consistent when considering a growing (instead of fixed) item population. [Appendix V]

more relevant (i.e., more accurately personalized) content to consumers. Second, understanding how data privacy practices affect RS performance *under different preference regimes* (i.e., in different recommendation settings that can exhibit varying degrees of user preference changes) makes the findings actionable across various domains. E.g., as discussed earlier, in application domains where user preferences are relatively stable, companies can prioritize opt-out-based data privacy practices; and in domains where user preferences can change dramatically, companies should prioritize deletion-based strategies. Third, the impact of data privacy practices on RS performance is a robust data-driven phenomenon that holds across different application scenarios and many traditional, popular recommendation algorithms. In other words, the insights from this study are not limited only to some highly specialized algorithmic solutions, and thus should enable system designers to adapt their RS to the newly emerging data privacy regulations in a wide variety of settings. At

the same time, when personal data is scarce due to data deletion, the choice of “fallback” recommendation strategies and exposure policies is highly consequential; e.g., heavier reliance on popularity-based fallback can lift short-run accuracy but reduce consumption diversity and disproportionately disadvantage niche products. Fourth, given the uncertain status of whether *inferences* (learned about users by machine learning models) are uniformly viewed as personal data, understanding the precise implications of *incremental learning* approaches (that can effectively utilize these inferences in the presence of data privacy practices) provides key guidance to system designers on when such approaches are advantageous to use. Fifth, the proposed parameterizable design based on revelation protection provides a nuanced and practical approach to implementing data privacy practices that are able to balance personalization and privacy considerations. In other words, our findings show that, under current regulations, there is still flexibility for companies to maintain their RS performance while providing data protection rights for users.

The results provide distinct, actionable guidance for consumers and policymakers as well. For consumers, regime-specific defaults can simplify privacy choices: in application domains where preferences are relatively stable, partial deletion that prioritizes protecting niche or rare activities (e.g., using the proposed revelation-protection parameterization with a moderate threshold) preserves recommendation quality while reducing preference revelation; in domains with more dynamic preferences, a full one-time deletion can jointly improve accuracy *and* reduce preference revelation by removing misaligned history. Finally, our findings can also be informative to policymakers when designing new regulations or revising existing ones, and could also lead policymakers towards further refinements of what should be considered as personal data and what data privacy practices should be provided and under what circumstances.

Future Work. This study points to a number of future research directions. First, as with any simulation-based approach, our study is based on (and is limited by) assumptions and modeling decisions that aim to represent key aspects of the application domain, its users, and RS. Future studies may investigate additional modeling choices that can further broaden our insights about data privacy practices and their impact on RS, for example, by incorporating more nuanced data privacy practices, more diverse user populations (say, in terms of their privacy-consciousness), different simulation time frames, more comprehensive user behavior (e.g., more complicated and subtle modeling of dynamic user preference change), etc. Second, our study is intentionally scoped to examine the effects of data privacy practices on technical RS performance (and the computational mechanisms through which it happens), measured by predictive accuracy, consumption diversity, and long-tail coverage, which are widely viewed as proximal mediators of eventual business impact. However, the magnitude of business outcome effects tends to be domain-, application-, and platform-specific and is best identified using large-scale randomized field experimentation (e.g., A/B tests) or natural experiments – while this is beyond the scope of this study, it is an important avenue for future work. Third, this study should be informative to future research on consumer data valuation and acquisition under data privacy regulations, as it provides an approach to explain the value of consumer data from the *algorithmic* performance perspective, complementing the more traditional economic and behavioral perspectives. Fourth, the current insights into how RS performance varies with different privacy practices (due to data loss) and with different preference dynamics (due to specifics of a given application domain) can guide the future designs of privacy-aware

RS. Additionally, while this study proposes revelation-protection parameterizations and characterizes their system-level performance trade-offs, a full HCI-theoretic design and evaluation of user-facing interfaces (e.g., comprehension of controls, communication of implications, accessibility/localization, and integration with the platform's privacy workflows) is beyond the present scope and constitutes important future work. Finally, the current insights can serve as a starting point to several follow-up studies, e.g., on exploring key aspects of privacy-related user behavior (that can impact RS performance) in specific application domains, such as understanding how/why/when users choose different data privacy practices or detecting users' preference changes.

References

- A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE security & privacy*, 3(1): 26–33, 2005.
- A. Acquisti, C. Taylor, and L. Wagman. The economics of privacy. *Journal of economic Literature*, 54(2):442–492, 2016.
- I. Adjerid, A. Acquisti, R. Telang, R. Padman, and J. Adler-Milstein. The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science*, 62(4):1042–1063, 2016.
- Adjust. What is app tracking transparency (att)?, 2021. <https://www.adjust.com/glossary/app-tracking-transparency/>.
- G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, 24(5):896–911, 2011.
- G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.
- T. Alashoor, M. Keil, H. J. Smith, and A. R. McConnell. Too tired and in too good of a mood to worry about privacy: Explaining the privacy paradox through the lens of effort level in information processing. *Information Systems Research*, 34(4):1415–1436, 2023.
- U. Alon and E. Yahav. On the bottleneck of graph neural networks and its practical implications. *arXiv preprint arXiv:2006.05205*, 2020.
- Amazon. Amazon.com privacy notice, 2023. Accessed 2024-02-07.
- Apple. Safari privacy overview. https://www.apple.com/safari/docs/Safari_White_Paper_Nov_2019.pdf, 2019. Accessed 2024-01-22.
- S. Athey, C. Catalini, and C. Tucker. The digital privacy paradox: Small money, small costs, small talk. Technical report, National Bureau of Economic Research, 2017.
- B. Auxier, L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner. Americans and privacy: Concerned, confused and feeling lack of control over their personal information. <https://www.pewresearch.org>, 2019. Accessed 2021-08-26.
- G. Bansal, D. Gefen, et al. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems*, 49(2):138–150, 2010.
- J. Bennett, S. Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, 2007.
- BestBuy. State privacy rights. <https://www.bestbuy.com/site/privacy-policy/state-privacy-rights>, 2023. Accessed 2023-07-14.
- CCPA. California consumer privacy act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5, 2018a. Accessed 2021-05-03.
- CCPA. California consumer privacy act. 2018b. URL https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB375. California Civil Code, Division 8, Part 4, Title 1.81.5, Section 1798.100 et seq.
- Y.-K. Che and J. Hörner. Recommender systems as mechanisms for social learning. *The Quarterly Journal of Economics*, 133(2):871–925, 2018.

- Z. Chen, Y. Wang, S. Zhang, H. Zhong, and L. Chen. Differentially private user-based collaborative filtering recommendation based on k-means clustering. *Expert Systems with Applications*, 168:114366, 2021.
- A. Collins, D. Tkaczyk, A. Aizawa, and J. Beel. A study of position bias in digital library recommender systems. *arXiv preprint arXiv:1802.06565*, 2018.
- CPA. U.s. state privacy laws in 2023. <https://www.lockelord.com/newsandevents/publications/2022/12/us-state-privacy-laws-2023>, 2022. Accessed 2023-07-10.
- CPIPL. China’s personal information protection law. https://www.dataguidance.com/sites/default/files/china_draft_personal_data_law.pdf, 2020. Accessed 2021-05-03.
- CTDPA. Connecticut data privacy act. <https://portal.ct.gov/ag/sections/privacy/the-connecticut-data-privacy-act>, 2022. Accessed 2024-05-03.
- A. D’Amour, H. Srinivasan, J. Atwood, P. Baljekar, D. Sculley, and Y. Halpern. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Conf. on Fairness, Accountability, and Transparency*, pages 525–534, 2020.
- M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1):143–177, 2004.
- B. J. Dietvorst, J. P. Simmons, and C. Massey. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of experimental psychology: General*, 144(1):114, 2015.
- DPDPA. Delaware personal data privacy act. <https://legis.delaware.gov/json/BillDetail/GenerateHtmlDocument?legislationId=140388&legislationTypeId=1&docTypeId=2&legislationName=HB154>, 2023. Accessed 2024-05-03.
- Ebay. User privacy notice, 2023. Accessed 2024-02-07.
- Etsy. United states regional privacy policy, 2023. Accessed 2024-02-07.
- European Union. Cookies, the gdpr, and the eprivacy directive, 2018. <https://gdpr.eu/cookies/>.
- C. Evans and A. Kasirzadeh. User tampering in reinforcement learning recommender systems. *arXiv preprint arXiv:2109.04083*, 2021.
- D. Fleder and K. Hosanagar. Blockbuster culture’s next rise or fall: The impact of recommender systems on sales diversity. *Management science*, 55(5):697–712, 2009.
- A. Friedman, B. P. Knijnenburg, K. Vanhecke, L. Martens, and S. Berkovsky. Privacy aspects of recommender systems. *Recommender systems handbook*, pages 649–688, 2015.
- GDPR. General data protection regulation. <https://gdpr-info.eu/>, 2018. Accessed 2021-05-03.
- S. Goldberg, G. Johnson, and S. Shriver. Regulating privacy online: The early impact of the gdpr on european web traffic & e-commerce outcomes. *Available at SSRN 3421731*, 2019.
- A. Goldfarb and C. E. Tucker. Privacy regulation and online advertising. *Management science*, 57(1):57–71, 2011.
- C. A. Gomez-Urbe and N. Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015.
- GDPL. General personal data protection law. <https://www.dlapiperdataprotection.com/index.html?t=law&c=BR>, 2018. Accessed 2021-05-03.
- R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.
- X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 549–558, 2016.
- X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.
- X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 639–648, 2020.
- H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang. Membership inference attacks on machine learning: A survey. *ACM Computing Surveys (CSUR)*, 54(11s):1–37, 2022.

- Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE international conference on data mining*, pages 263–272. Ieee, 2008.
- D. Jannach, M. Ludewig, and L. Lerche. Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. *User Modeling and User-Adapted Interaction*, 27:351–392, 2017.
- J. Jia, G. Z. Jin, and L. Wagman. The short-run effects of gdpr on technology venture investment. Technical report, National Bureau of Economic Research, 2018.
- C. C. Johnson. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27(78):1–9, 2014.
- W.-C. Kang and J. McAuley. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pages 197–206. IEEE, 2018.
- A. A. Khaliq, A. Anjum, A. B. Ajmal, J. L. Webber, A. Mehbodniya, and S. Khan. A secure and privacy preserved parking recommender system using elliptic curve cryptography and local differential privacy. *IEEE Access*, 10: 56410–56426, 2022.
- K. Klöckner, N. Wirschum, and A. Jameson. Depth-and breadth-first processing of search result lists. In *CHI’04 extended abstracts on Human factors in computing systems*, pages 1539–1539, 2004.
- A. Kobsa. Privacy-enhanced web personalization. In *The adaptive web: Methods and strategies of web personalization*, pages 628–670. Springer, 2007.
- S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, 64:122–134, 2017.
- J. A. Konstan and J. Riedl. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction*, 22:101–123, 2012.
- J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87, 1997.
- Y. Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 447–456, 2009.
- E. Laziuk. ios 14.5 opt-in rate, 2021. <https://www.flurry.com/blog/ios-14-5-opt-in-rate-att-restricted-app-tracking-transparency-worldwide-us-daily-latest-update>.
- K. Leswing. Apple’s ad privacy change impact shows the power it wields over other industries, 2021. <https://www.cnn.com/2021/11/13/apples-privacy-changes-show-the-power-it-holds-over-other-industries.html>.
- T. Li, L. Song, and C. Fragouli. Federated recommendation system via differential privacy. In *2020 IEEE international symposium on information theory (ISIT)*, pages 2592–2597. IEEE, 2020.
- X. Li, J. Grahl, and O. Hinz. How do recommender systems lead to consumer purchases? a causal mediation analysis of a field experiment. *Information Systems Research*, 33(2):620–637, 2022.
- Y. Li. The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns. *Decision support systems*, 57:343–354, 2014.
- T. Lin and A. Strulov-Shlain. Choice architecture, privacy valuations, and selection bias in consumer data. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2023-58), 2023.
- Y. Liu, Y. Su, A.-A. Liu, B. Schiele, and Q. Sun. Mnemonics training: Multi-class incremental learning without forgetting. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- J. M. Logg, J. A. Minson, and D. A. Moore. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103, 2019.
- T. Lohrenz, K. McCabe, C. F. Camerer, and P. R. Montague. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences*, 104(22):9493–9498, 2007.
- M. Ludewig and D. Jannach. Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, 28(4-5):331–390, 2018.
- McKinsey. How retailers can keep up with consumers. [https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers#/,](https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers#/) 2024. Accessed: 2024-05-16.
- L. Minto, M. Haller, B. Livshits, and H. Haddadi. Stronger privacy for federated collaborative filtering with implicit feedback. In *Proceedings of the 15th ACM Conference on Recommender Systems*, pages 342–350, 2021.

- W. W. Moe. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, 13(1-2):29–39, 2003.
- I. Mori. The trust opportunity: Exploring consumer attitudes to the internet of things. <https://www.internetsociety.org/>, 2019. Accessed 2021-08-26.
- A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE, 2008.
- K. F. Pereira. Measuring forgetting in stream-based recommender systems. 2022.
- D. Rafailidis and A. Nanopoulos. Modeling users preference dynamics and side information in recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(6):782–792, 2015.
- S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback, 2012.
- F. Ricci, L. Rokach, and B. Shapira, editors. *Recommender Systems Handbook*. Springer, 3 edition, 2022. doi: 10.1007/978-1-0716-2197-4. URL <https://link.springer.com/book/10.1007/978-1-0716-2197-4>.
- B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *The 10th international Conf. on World Wide Web*, pages 285–295, 2001.
- A. Shen and A. D. Ball. Preference stability belief as a determinant of response to personalized recommendations. *Journal of Consumer Behaviour*, 10(2):71–79, 2011.
- C. Song and A. Raghunathan. Information leakage in embedding models. In *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, pages 377–390, 2020.
- W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang. Session-based social recommendation via dynamic graph attention networks. In *Proc. of the ACM Intl. Conf. on Web Search and Data Mining*, pages 555–563, 2019.
- T. Sun, Z. Yuan, C. Li, K. Zhang, and J. Xu. The value of personal data in internet commerce: A high-stake field experiment on data regulation policy. *SSRN 3566758*, 2020.
- S. Swarnkar and N. Kalla. A study on preferential shift of consumer choices in their buying behavior towards the e-commerce. *NeuroQuantology*, 19(6):253, 2021.
- G. Takács, I. Pilászy, and D. Tikk. Applications of the conjugate gradient method for implicit feedback collaborative filtering. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 297–300, 2011.
- R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.
- UCPA. Utah consumer privacy act. <https://dcp.utah.gov/ucpa/#:~:text=The%20UCPA%20provides%20Utah%20consumers,and%20request%20it%20be%20deleted.>, 2023. Accessed 2024-05-03.
- VCDPA. Consumer data protection act. <https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/>, 2023. Accessed 2023-07-10.
- Q. Wang, H. Yin, T. Chen, J. Yu, A. Zhou, and X. Zhang. Fast-adapting and privacy-preserving federated recommender system. *The VLDB Journal*, 31(5):877–896, 2022.
- WPA. Washington privacy act. <https://app.leg.wa.gov/billssummary?BillNumber=5062&Year=2021&Initiative=False>, 2021. Accessed 2021-05-03.
- T. Yao, X. Yi, D. Z. Cheng, F. Yu, T. Chen, A. Menon, L. Hong, E. H. Chi, S. Tjoa, J. Kang, et al. Self-supervised learning for large-scale item recommendations. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 4321–4330, 2021.
- K. Yu, S. Berkovsky, R. Taib, D. Conway, J. Zhou, and F. Chen. User trust dynamics: An investigation driven by differences in system performance. In *Proceedings of the 22nd international conference on intelligent user interfaces*, pages 307–317, 2017.
- K. Yu, S. Berkovsky, D. Conway, R. Taib, J. Zhou, and F. Chen. Do i trust a machine? differences in user trust based on system performance. In *Human and machine learning: Visible, explainable, trustworthy and transparent*, pages 245–264. Springer, 2018.
- Z. Yuan, A. Y. Chen, Y. Wang, and T. Sun. How recommendation affects customer search: A field experiment. *Information Systems Research*, 2024.
- J. Zhang, G. Adomavicius, A. Gupta, and W. Ketter. Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Information Systems Research*, 31(1): 76–101, 2020.

- Z. Zhang. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee, 2018.
- M. Zhou, J. Zhang, and G. Adomavicius. Longitudinal impact of preference biases on recommender systems’ performance. *Information Systems Research*, 2023.

A Examples of Data Privacy Practices from Online Platforms

Figure A11: BestBuy Provides Consumer Portal for Manually Deleting Past Data

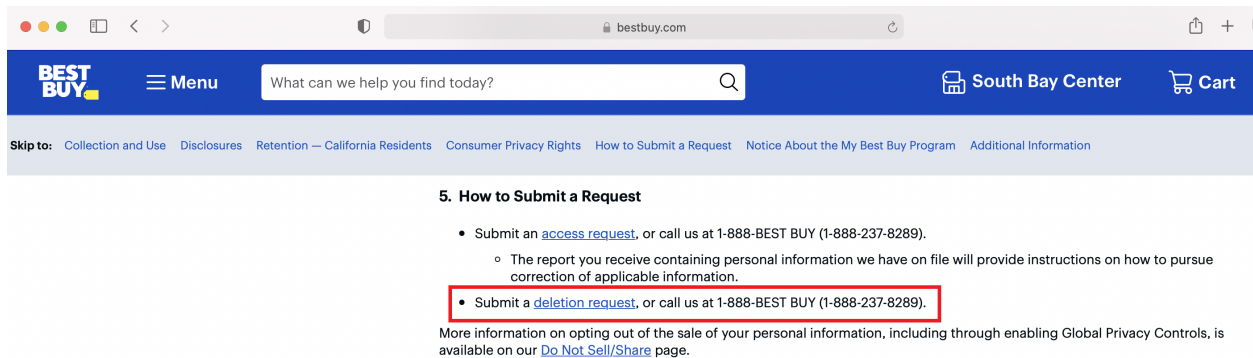


Figure A12: Youtube Provides Consumer Portal for Auto-Deleting Past Data

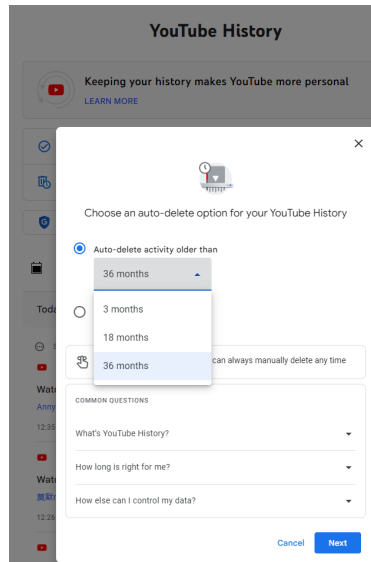
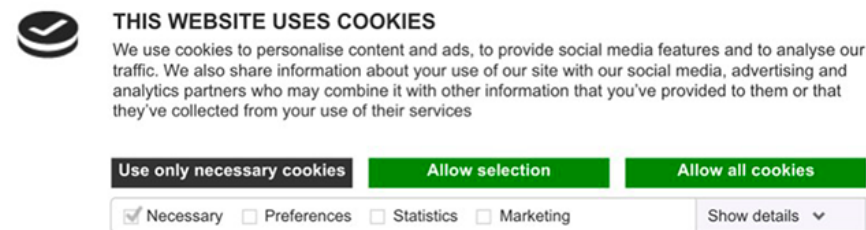


Figure A13: Cookie Consent Banner



B Main Simulation Components and Notation

Table B4: Key Notation for the Simulation Framework

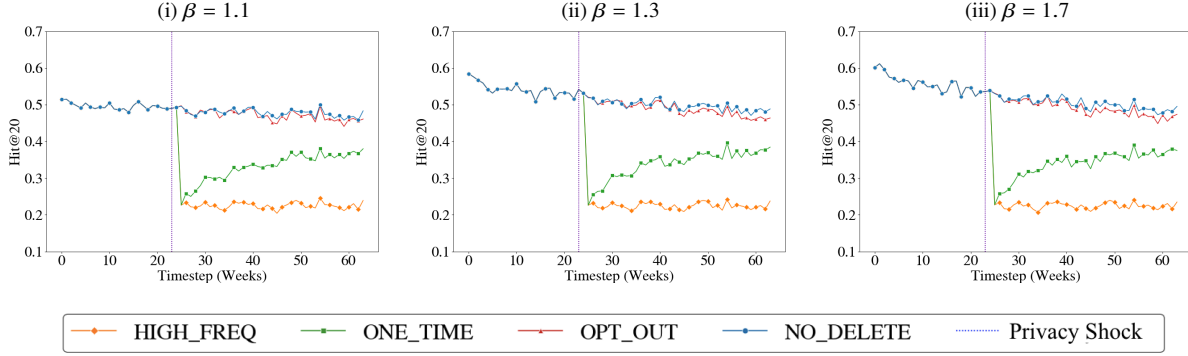
Aspect	Notations and Explanations
A. Item population	
Item set	I : elements of the set are individual items indexed by i ($i \in I$).
Item content	e_i : A latent vector of item i 's ($i \in I$) content features
B. User population	
User set	U : elements of the set are individual users indexed by u ($u \in U$).
User preferences	e_u or $\hat{e}_u$: latent vector that represents user u 's ($u \in U$) stable or dynamic preference profile.
Degree of user preference change	α : the degree of change in user preference profile, with 0 indicating original (unchanged) profile and 1 indicating a complete change into a different profile.
User needs	$PrefList_u$: User u 's list of all items ranked by the inherent user preferences.
User reliance on recommendations	w : the probability of user's consumption from $RecList$ during a given visit; correspondingly, $(1 - w)$ indicates user's consumption probability from $PrefList_u$.
Recommendation list exploration	β : controls how far down the recommendation list users are likely to explore before choosing an item.
User consumption frequency	Δ_{ut} : representation of length of time between user u 's consumption at t and the next consumption. Also, next scheduled next consumption for user u is referred to as $UserNextCons_u$ value.
User consumption data	$ConsData_u^t$: all historical consumption data of user u up to (and including at) time t .
User privacy request	$DataRequest_u$: the enactment of a certain data privacy practice by the user u ; for example, delete all personal data under ONE_TIME.
C. Recommender engine	
Recommendation generation	$GenerateRecs$: recommendation generator(s) used to train the RS and produce a list of recommendations for each user for the next time period.
Performance measurement	$Evaluate$: performance metrics, for measuring different aspects of recommender systems performance, for example, accuracy, diversity, etc.
Training Data	$TrainData_u^t$ represents user-item consumption history that is made available to RS, which is often a subset of all consumption data $ConsData_u^t$ due to privacy practices.
System Recommendations	$RecList_u^t$ represents the list of recommended items ranked by the system's predicted preferences of user u at time t
D. Scenario Change	
Privacy Shock	$T_{DataReq}$ is the time when users can start requesting data privacy actions.
Preference Transition	$T_{PrefChange}$ is the time when users' preferences change.

C Recommendation List Exploration

As discussed in Section 4.1, parameter β controls the extent to which each user explores the recommended item list before making a selection for item consumption. We vary the value of β with 1.1, 1.3 (the default choice), and 1.7, and the results under stable user preferences are presented in Figure C14. These three values of β approximately represent scenarios where the user consumes from top-20, top-10, and top-5 recommendations, respectively; for example, the top-1 item has 10.7%, 23.1%, and 41.2% chance of being consumed with these three values of β , respectively. Notably, the value of β does not affect the key qualitative patterns of main results, while the overall RS performance increases for larger β values. Since larger β s concentrate

the probability mass more towards the top-ranked items, the users' consumption behavior becomes more predictable and thus the system performance is improved. The consistency of main results with different β values verifies the robustness of the simulation framework and of the key insights.

Figure C14: Recommendation List Exploration, Static User Preferences, Affected Users, Hit@20, BPR



Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

D Main Experiment Results with Additional Accuracy Metrics: MRR@20 and NDCG@20

Mean Reciprocal Rank of top- K recommended items (MRR@ K) evaluates the ranking quality by considering the position of the first relevant item in the recommendation list. It is defined as:

$$\text{MRR@K} = \frac{1}{|U|} \sum_{u \in U} \frac{1}{\text{rank}_u} \quad (7)$$

where rank_u is the position of the first relevant item in user u 's recommendation list RecList_u^t within the top- K items. If no relevant items are found within top- K , rank_u is set to infinity, resulting in a reciprocal rank of zero.

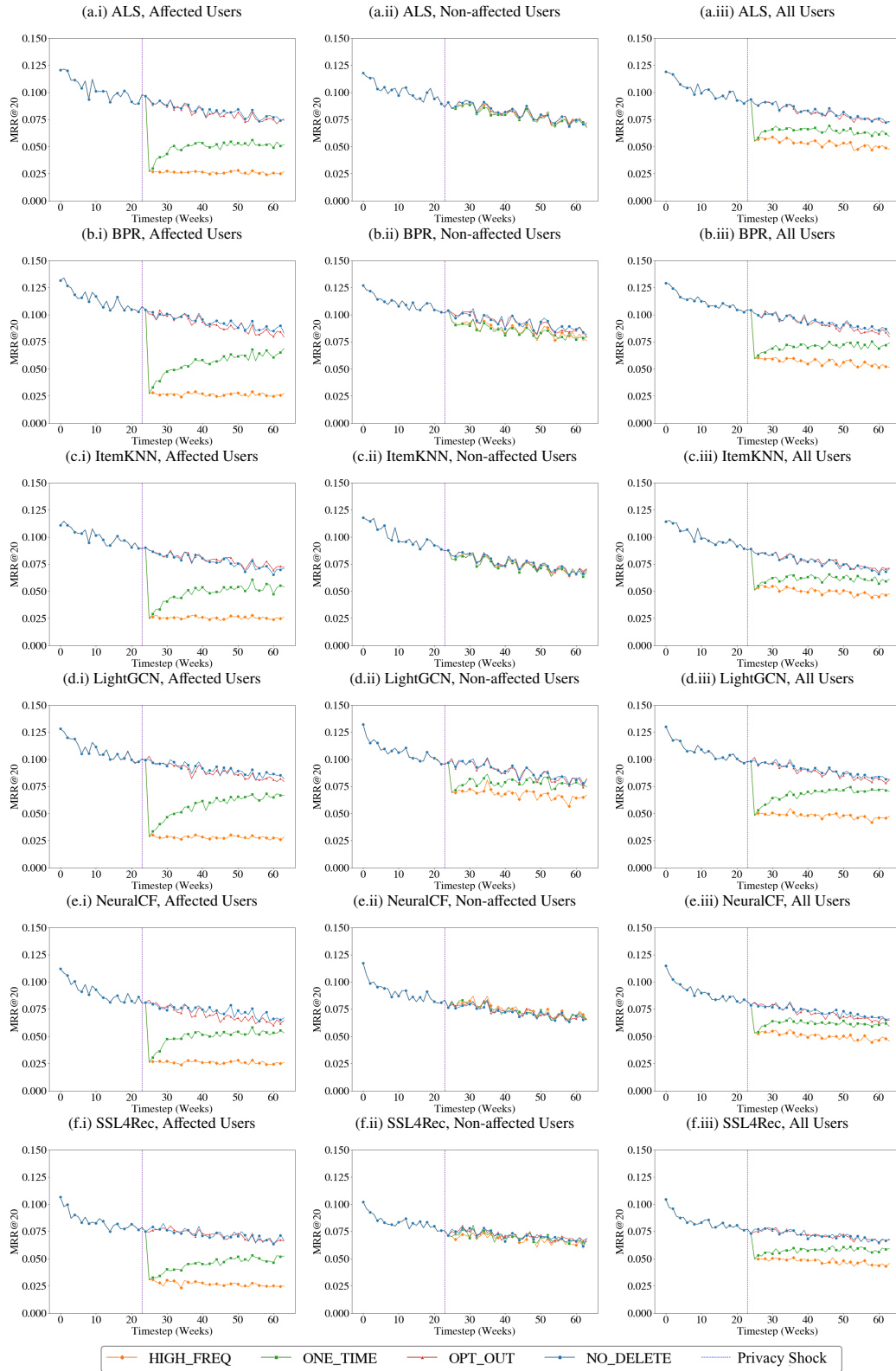
Normalized Discounted Cumulative Gain of top- K recommended items (NDCG@ K) measures ranking quality by considering both the position and relevance of items in the recommendation list:

$$\text{NDCG@K} = \frac{1}{|U|} \sum_{u \in U} \frac{\sum_{r=1}^K \frac{\text{rel}_{u,r}}{\log_2(r+1)}}{\sum_{r=1}^K \frac{1}{\log_2(r+1)}} \quad (8)$$

where $\text{rel}_{u,r} = 1$ if the item at rank r in RecList_u^t is relevant for user u , and 0 otherwise. The denominator represents the ideal score when all relevant items are ranked at the top.

These metrics provide insights into the accuracy of RS in ranking relevant items at the top positions. We provide main experiment results in terms of MRR@20 and NDCG@20 below. They are consistent with the results (in terms of Hit@20) presented in the main paper.

Figure D15: Impact of Data Privacy Practices with Stable User Preferences, MRR@20



Note that across all figures, 50% of users are affected, and all users' reliance on recommendations is 20%.

Figure D16: Impact of Data Privacy Practices with Dynamic User Preferences, MRR@20

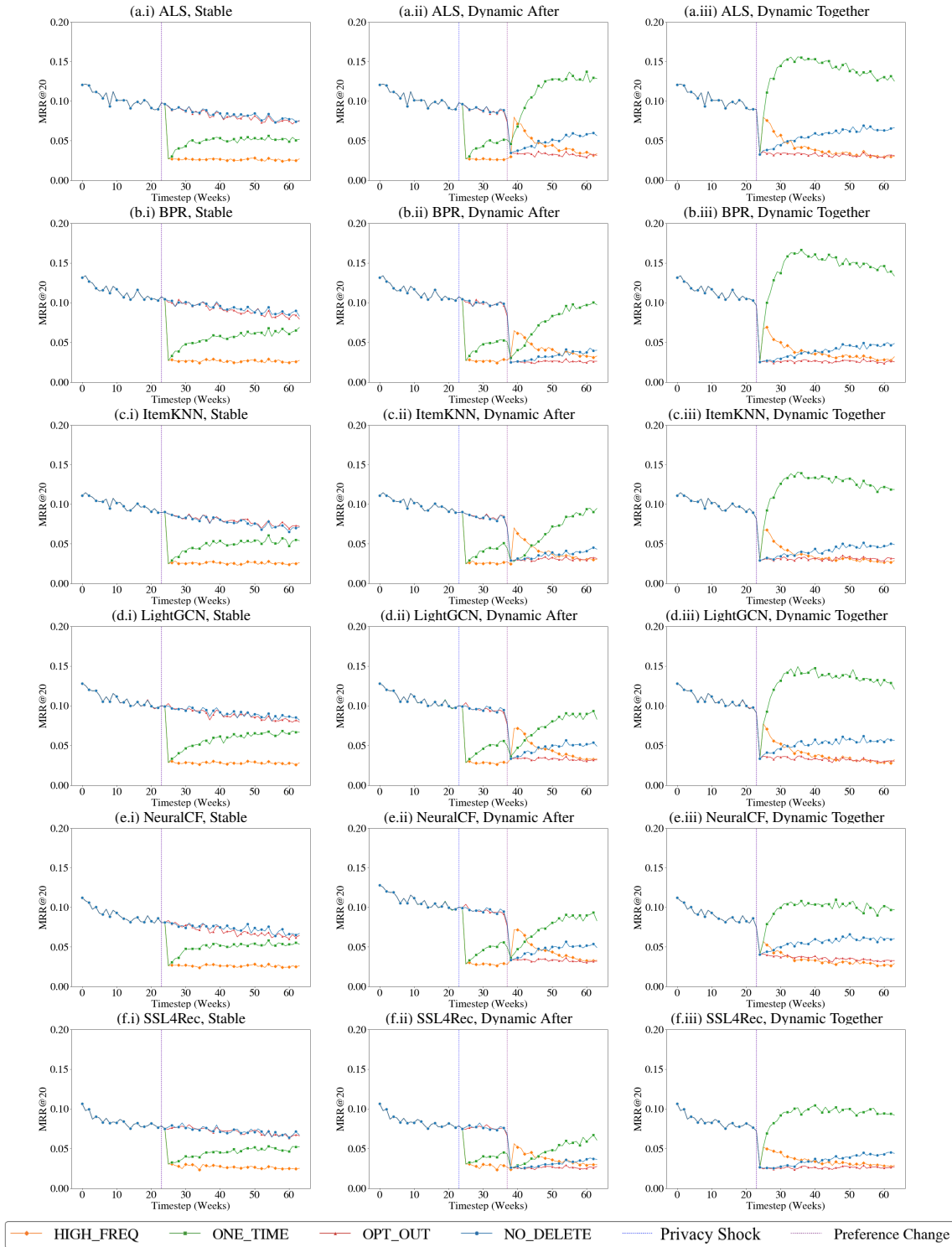
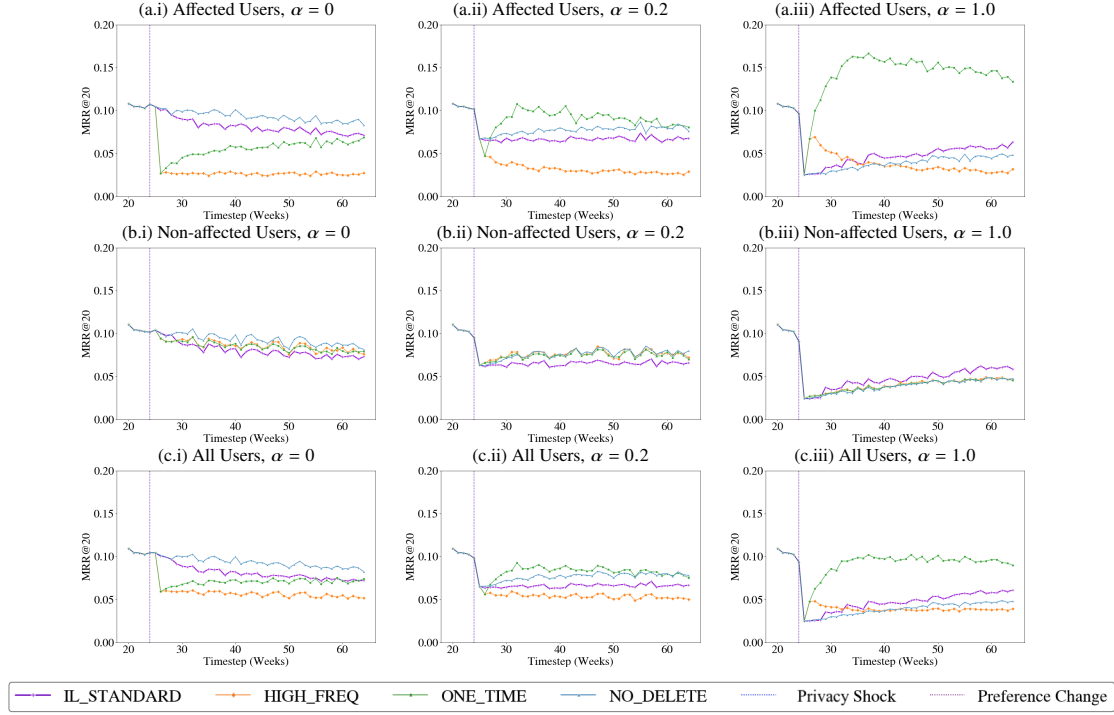
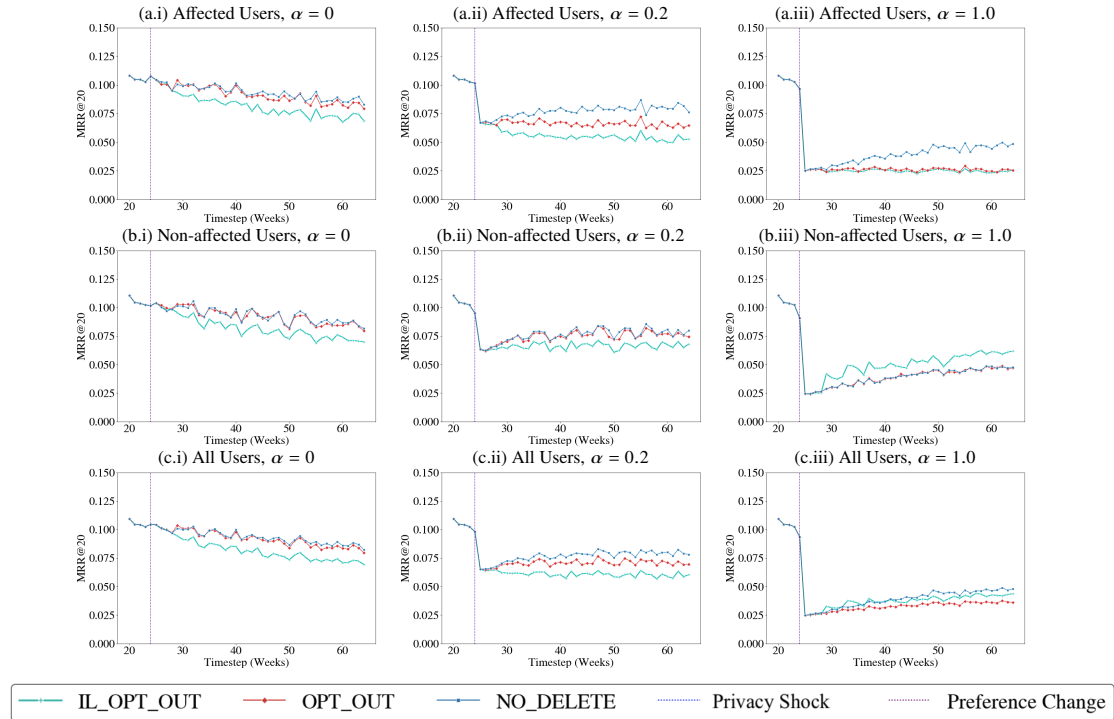


Figure D17: Incremental and Non-incremental Learning under Deleting Practices, MRR@20



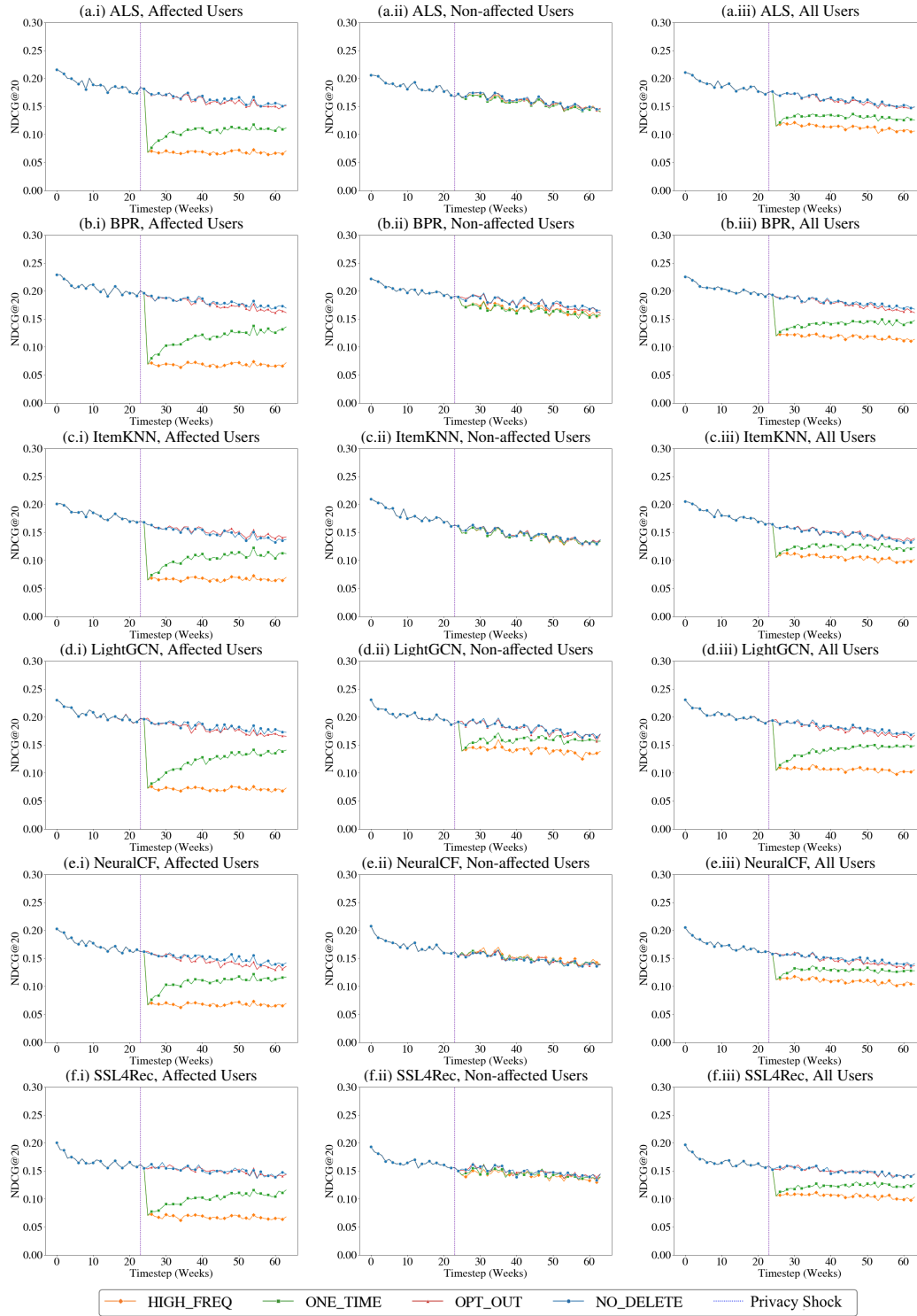
The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure D18: Incremental and Non-incremental Learning under Opting Out Practice, MRR@20



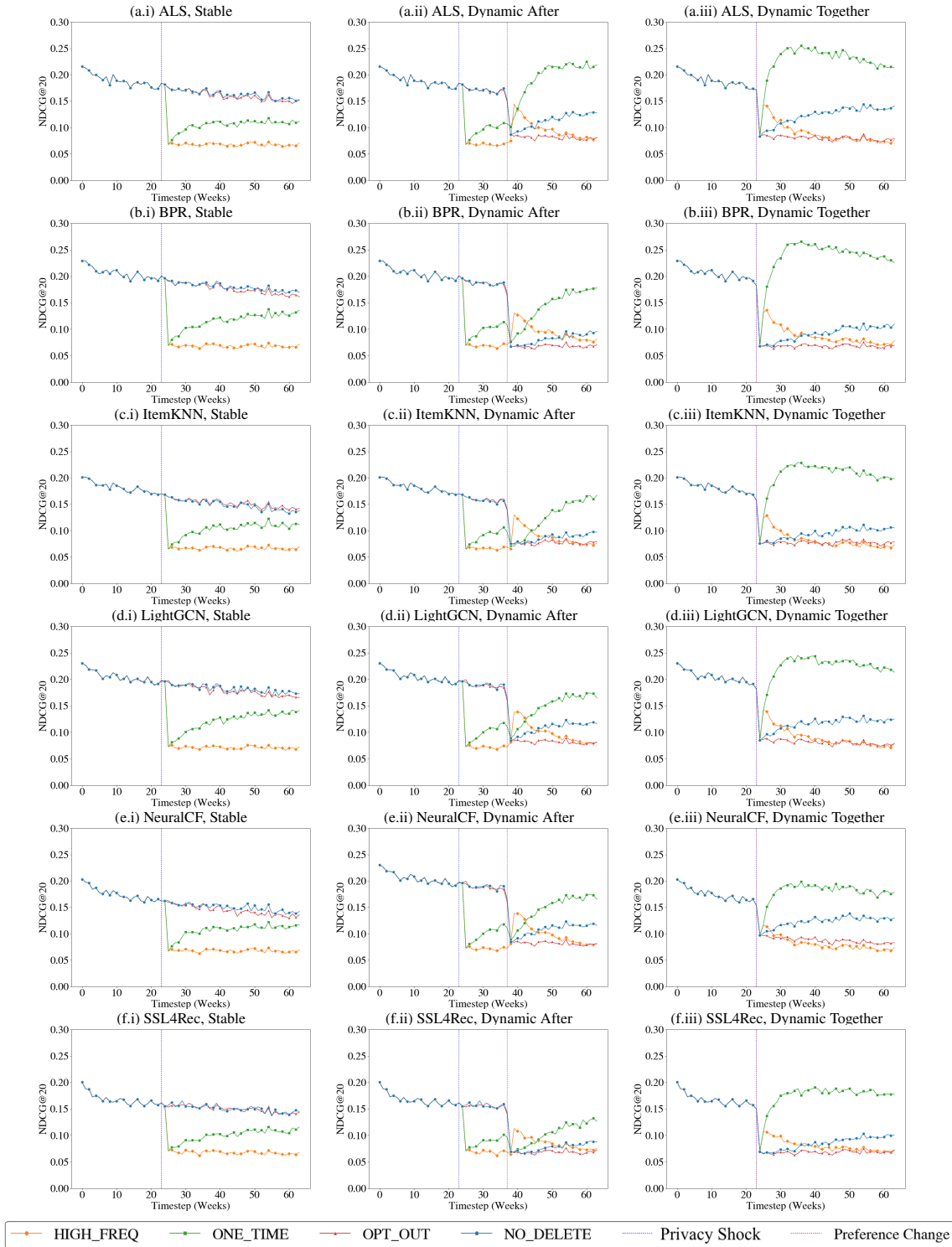
The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure D19: Impact of Data Privacy Practices with Stable User Preferences, NDCG@20



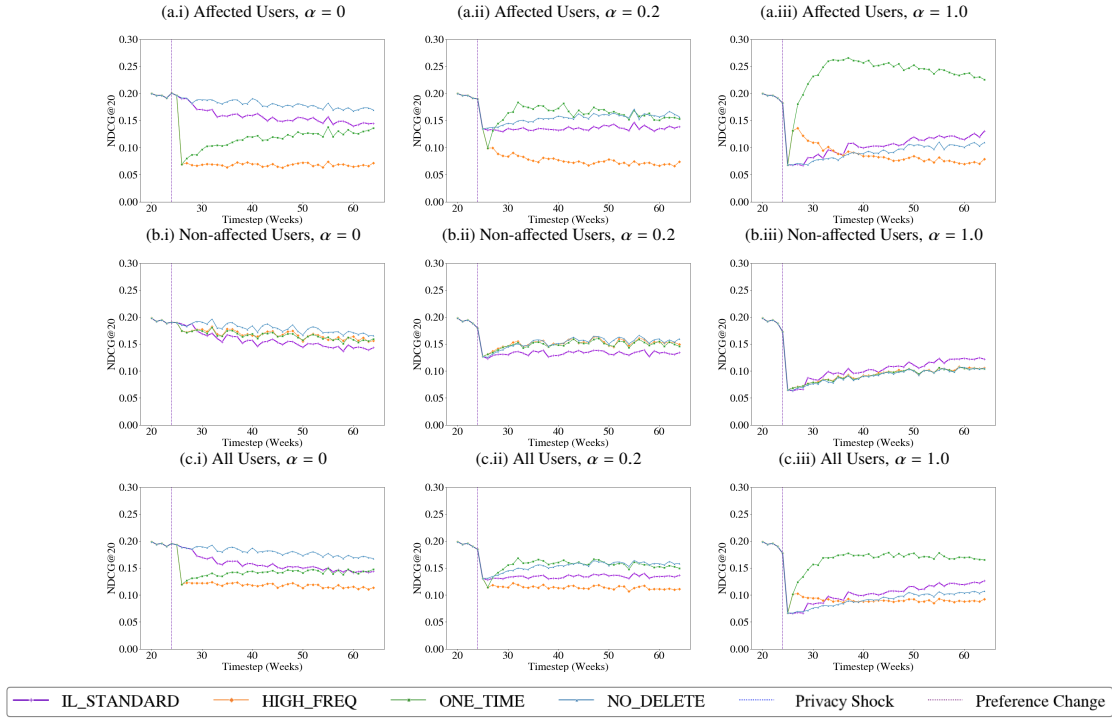
Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

Figure D20: Impact of Data Privacy Practices with Dynamic User Preferences, NDCG@20



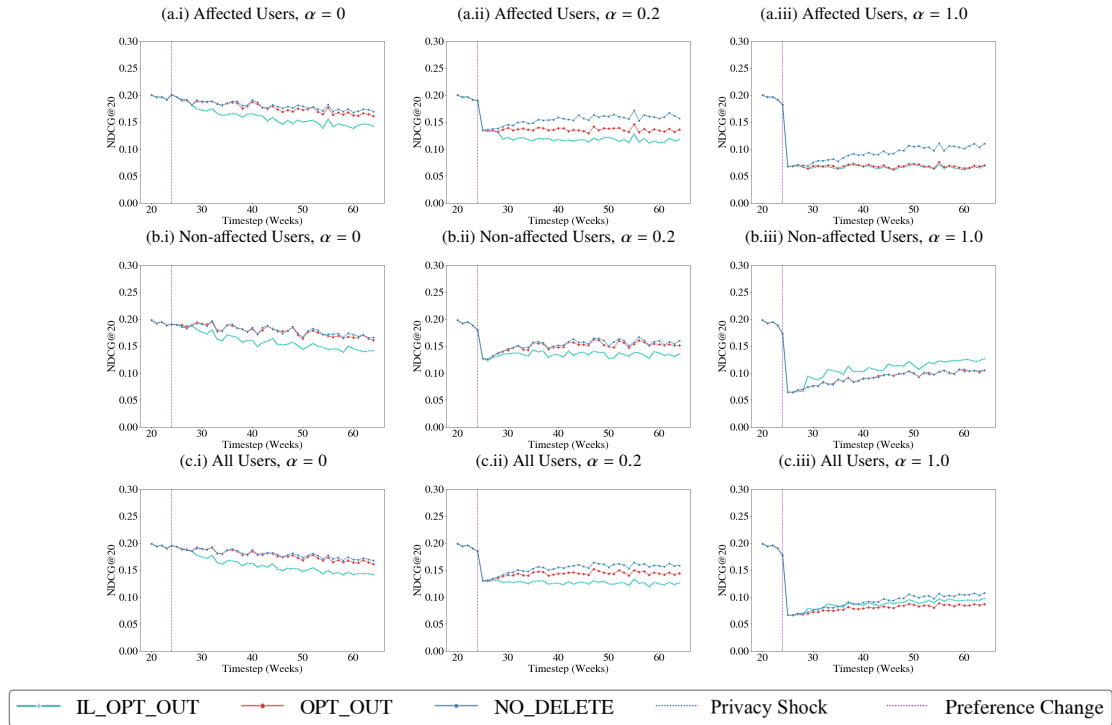
Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

Figure D21: Incremental and Non-incremental Learning under Deleting Practices, NDCG@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure D22: Incremental and Non-incremental Learning under Opting Out Practice, NDCG@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

E Main Experiment Results with Additional Seed Datasets

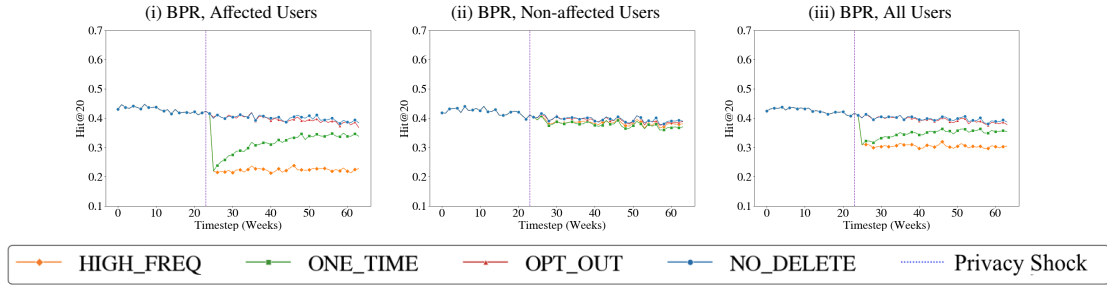
In this section, we present main experiment results with additional seed datasets *Amazon Book* and *Retailrocket*. The summary statistics of all datasets (i.e., the main dataset used in the paper and two additional seed datasets) are shown in Table E5.

Table E5: Summary Statistics of Seed Datasets

Dataset	Domain (Interactions)	Sparsity Level	User/Product Ratio
Multi-Channel Retailer	Retail (Purchases)	1.438×10^{-4}	18,938/40,716 (0.465)
Retailrocket	Retail (Views)	8.079×10^{-6}	1,404,179/234,838 (5.98)
Amazon Book	Books (Purchases)	4.005×10^{-5}	603,668/367,982 (1.64)

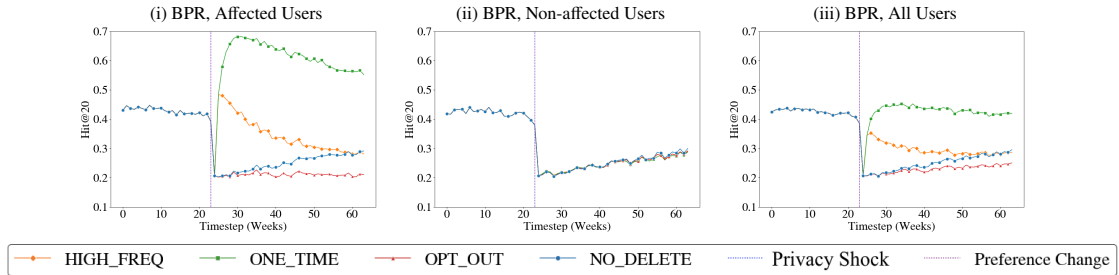
E.1 Main Experiment Results with *Amazon Book*

Figure E23: Impact of Data Privacy Practices with Stable User Preferences, Amazon Book, Hit@20



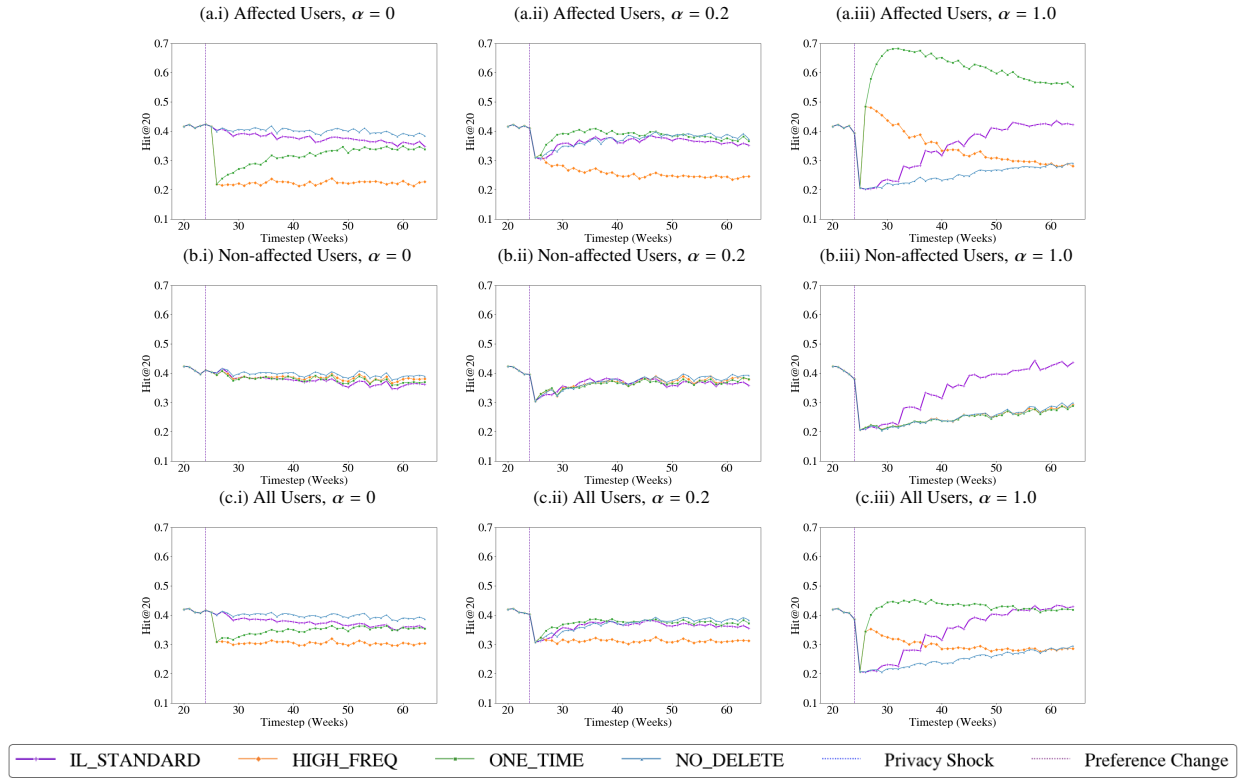
Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

Figure E24: Impact of Data Privacy Practices with Dynamic User Preferences, Amazon Book, Hit@20



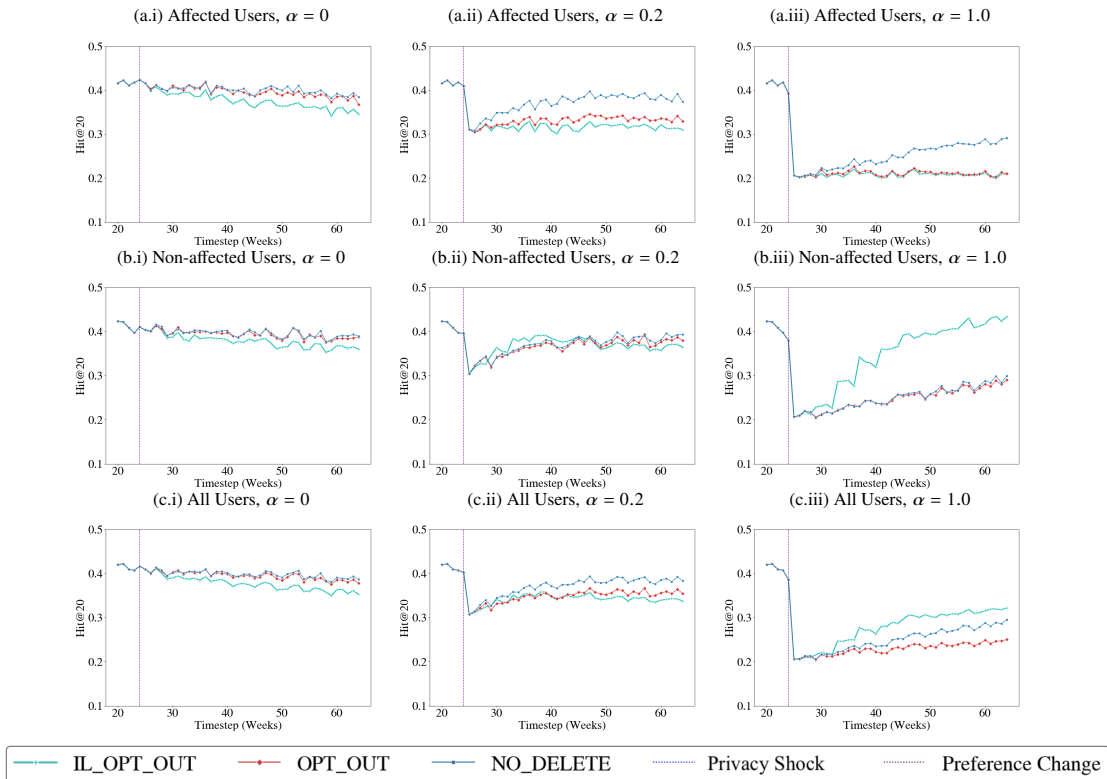
Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

Figure E25: Incremental and Non-incremental Learning under Deleting Practices, Amazon Book, Hit@20



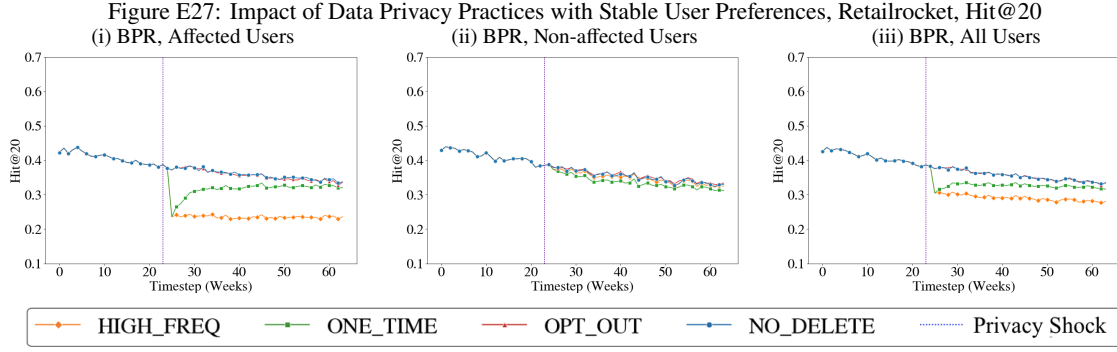
The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure E26: Incremental and Non-incremental Learning under Opting Out Practice, Amazon Book, Hit@20

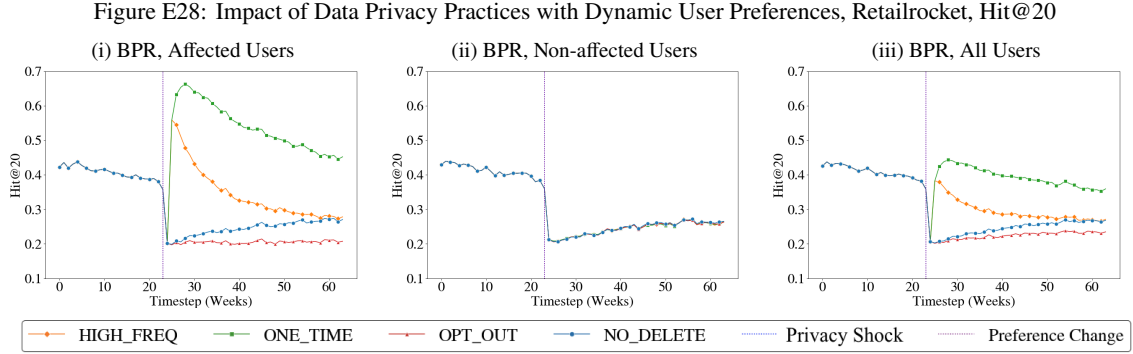


The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

E.2 Main Experiment Results with *Retailrocket*

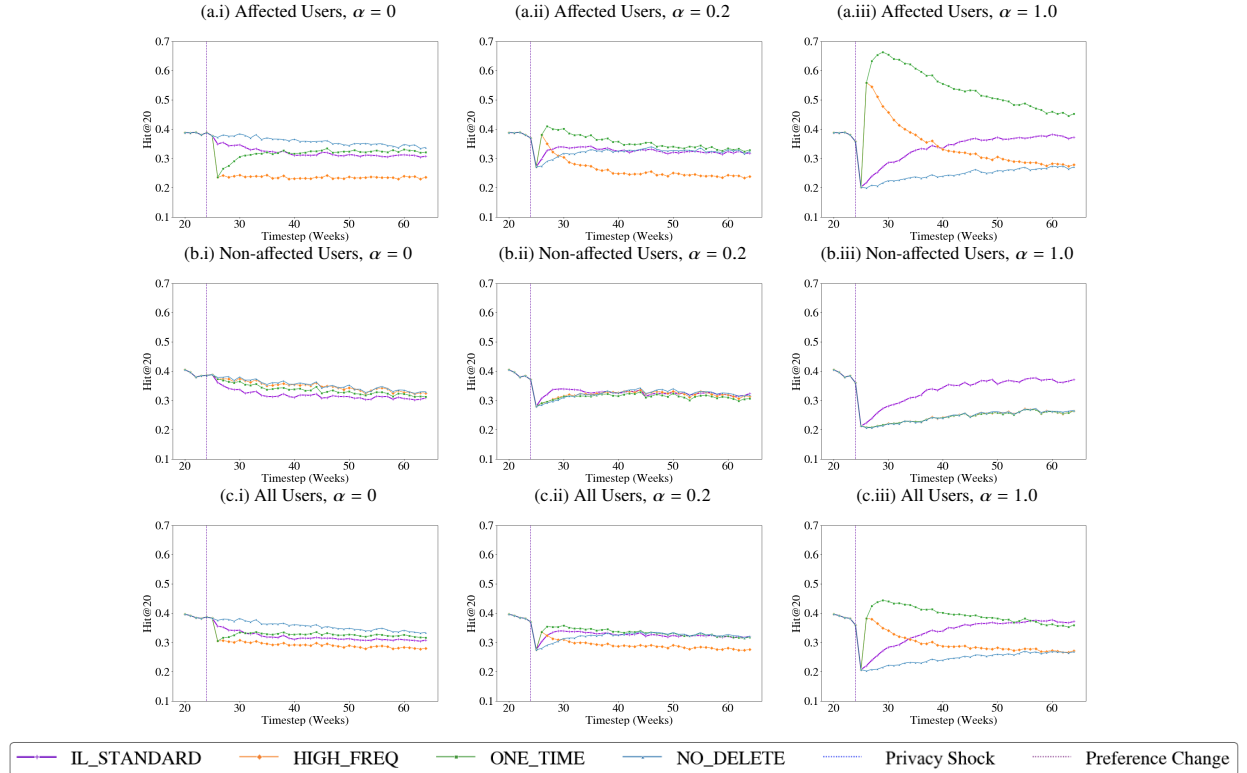


Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.



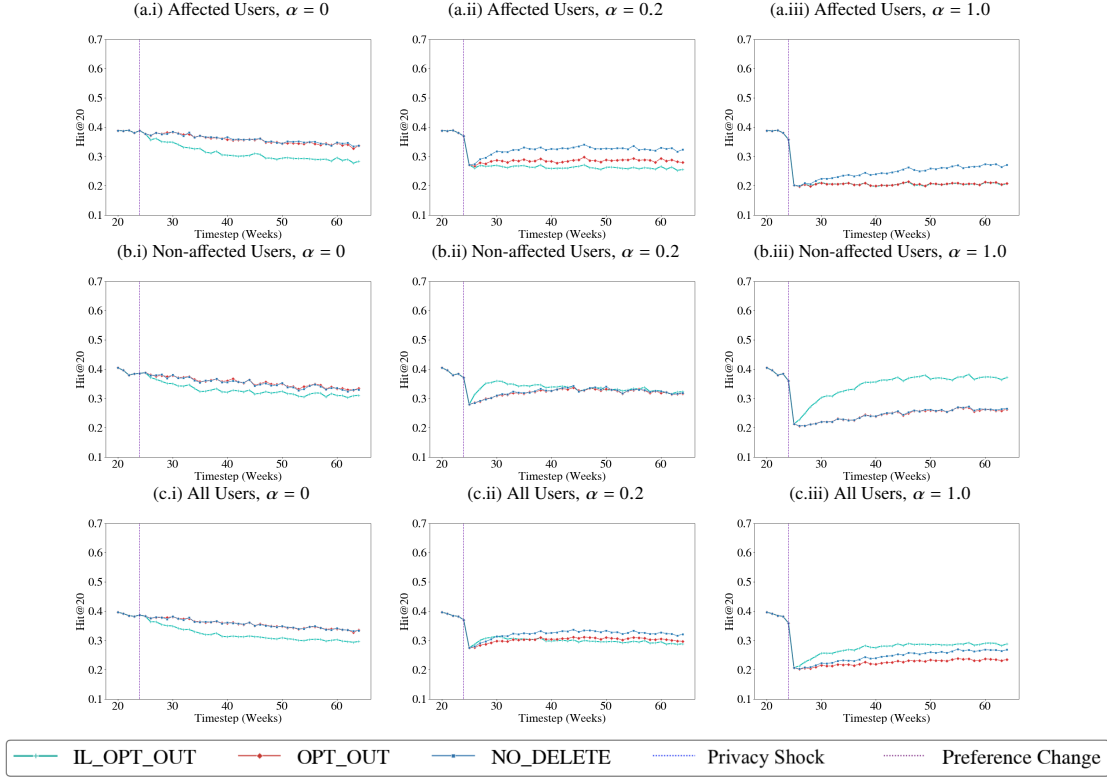
Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

Figure E29: Incremental and Non-incremental Learning under Deleting Practices, Retailrocket, Hit@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure E30: Incremental and Non-incremental Learning under Opting Out Practice, Retailrocket, Hit@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

F Benchmarking/Validation Results for Recommender System Algorithms

To ensure validity and correctness of our simulation framework, we benchmarked our implementations of the recommender system (RS) algorithms, specifically Neural Collaborative Filtering (NeuralCF), LightGCN, and SSL4Rec. We adopted the official implementations (i.e., without any modifications) of Bayesian Personalized Ranking (BPR), Alternating Least Squares (ALS), and ItemKNN models directly from the LensKit recommender package;¹⁹ and, thus, did not include them in this benchmarking exercise.

We conducted the benchmarking and compared our results with the results reported in the respective original papers (using publicly available datasets as well as the metrics mentioned in the original papers). Tables F6, F7, and F8 summarize these comparisons. Our implementations align closely with the architectures and procedures described in the original works. Although our benchmarking results do not match the reported values precisely (due to some inherent random elements in the model training and evaluation processes and minor implementation variations), the performance numbers are highly comparable.

¹⁹LensKit recommender package: <http://lenskit.org/>

Table F6: Benchmarking Results for NeuralCF

	Hit@10	NDCG@10
Reported	0.701	0.425
Our Implementation	0.687	0.411

Benchmarking is done on the MovieLens dataset. For benchmarking purposes, our implementation of NeuralCF has 32 hidden factors and 3 hidden layers (# of weights: 128, 64, 32) without pre-training. We follow the same finetuning process as described in Section 4.1 of the original NeuralCF paper. Reported results are from Table 2 of the original NeuralCF paper. Results of our implementation represent an average across 5 runs (using different random seeds).

Table F7: Benchmarking Results for LightGCN

	Recall@20	NDCG@20
Reported	0.0411	0.0315
Our Implementation	0.0408	0.0313

Benchmarking is done on the Amazon Book dataset. For benchmarking purposes, our implementation of LightGCN has 2 hidden layers (# of weights: 64) with a learning rate of 0.001 and default mini-batch size of 1024. We follow the same finetuning process as described in Section 4.1.2 of the original LightGCN paper. Reported results are from Table 3 of the original LightGCN paper. Results of our implementation represent an average across 5 runs (using different random seeds).

Table F8: Benchmarking Results for SSL4Rec

	MAP@10	Recall@20
Reported	0.0171	0.0573
Our Implementation	0.0162	0.0568

Benchmarking is done on the Wikipedia dataset. For benchmarking purposes, our implementation of SSL4Rec aligns with the baseline architecture described in Section 4.2 of the original SSL4Rec paper, and we follow the same hyper-parameter search process as described in that same section. Reported results are from Table 2 of the original SSL4Rec paper. Results of our implementation represent an average across 5 runs (using different random seeds).

G Hyper-Parameter Tuning for Recommender System Algorithms

The six recommendation algorithms examined in this study cover a wide range of families of recommender system models. ALS and BPR are two canonical representations of the matrix factorization models. ItemKNN is a notable example of the traditional neighborhood-based collaborative filtering methods. NeuralCF, LightGCN, and SSL4Rec represent the family of recommendation algorithms based on modern neural network architectures.

The hyper-parameters of these models are tuned on historical data *ConsData*⁰. For the ALS model, we did a grid search over three features: the number of latent factors {8, 16, 32}, the number of iterations {10, 20, 50}, the regularization factor {0.1, 0.2, 0.4, 0.8}, and the weight on positive instances {10, 20, 40}. For the BPR model, we did a grid-search over: the number of latent factors {8, 16, 32}, the number of iterations {10, 20, 50}, the regularization factor {0.1, 0.2, 0.4, 0.8}, and the learning rate {0.01, 0.05, 0.1}. For the ItemKNN model, we did a grid-search over: the maximum number of neighbors {10, 20, 30, 50, 80}, the minimum number of neighbors {1, 5, 10}, the minimum similarity to be considered a neighbor $\{1 \times 10^{-6}, 1 \times 10^{-4}, 1 \times 10^{-2}, 1 \times 10^{-1}\}$,

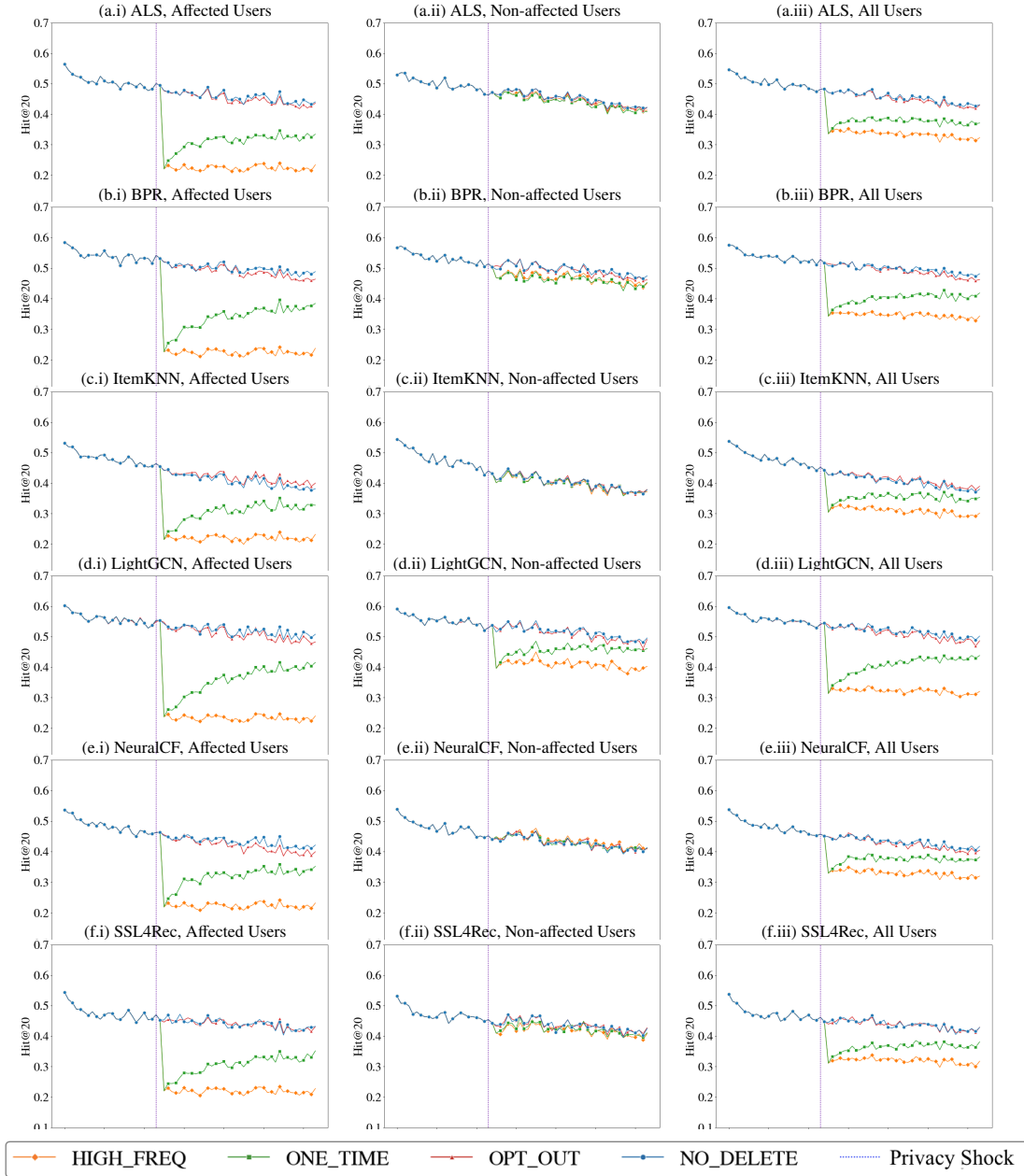
and the aggregation scheme using either the simple sum or the weighted sum. For the LightGCN model, we searched over: the number of latent features $\{8, 16, 32\}$, the learning rate $\{0.001, 0.005, 0.01, 0.05\}$, and the dropout rate $\{0.0, 0.2, 0.4, 0.6\}$. For the NeuralCF model, we tested a wide neural architecture and a deep neural architecture for the Multi-layer Perceptron (MLP) part of the model. Specifically, for a wide neural architecture, we have one layer of MLP with 64 latent features. For a deeper neural architecture, we have 3 layers of MLPs with 32, 16, and 8 latent features respectively. We also tune the dropout for the MLP layers using a set of values $\{0.2, 0.4, 0.6\}$. For the Generalized Matrix Factorization (GMF) layer, we searched over a number of latent features $\{8, 16, 32\}$ and the learning rate $\{0.0001, 0.0005, 0.001, 0.005\}$. We trained the best configuration of LightGCN and NeuralCF to convergence with an Adam optimizer (Zhang 2018). For the SSL4Rec model, we conducted a grid search over four key hyper-parameters. Specifically, we tuned the dimensionality of the latent embeddings over the set $\{64, 128, 256\}$, as this parameter governs the expressiveness of the user and item representations passed to the neural towers. We varied the dropout rate applied to item embeddings over the set $\{0.0, 0.1, 0.2, 0.4\}$ to assess its effect on the contrastive learning regularization. For the learning rate, we evaluated values in the set $\{0.0005, 0.001, 0.005, 0.01\}$, as this controls the step size in the parameter updates via the Adam optimizer. Finally, we explored different weights on the contrastive learning loss $\alpha \in \{0.05, 0.1, 0.2, 0.5\}$, which governs the trade-off between the recommendation loss and the self-supervised auxiliary loss. The temperature parameter τ used in both the recommendation loss and the contrastive loss was tuned over $\{0.05, 0.1, 0.2\}$. We trained the final model configuration to convergence using the Adam optimizer.

As a result of the fine-tuning process, we select the configuration of the model that yields best *Hit@20*, and fix the configuration throughout the simulation period. For ALS, the best configuration is: the number of latent factors, 16; the number of iterations, 20; the regularization factor, 0.1; the weight on positive instances, 40. For BPR, the best configuration is: the number of latent factors, 32; the number of iterations, 100; the regularization factor, 0.01. For ItemKNN, the best configuration is: the maximum number of neighbors, 50; the minimum number of neighbors, 1; the minimum similarity to be considered a neighbor, 1×10^{-6} ; and the weighted sum as the aggregation scheme. For LightGCN, the best configuration is: the number of latent factors, 16; the number of latent factors of the GMF layer, 0.0; the dropout rate, 0.0. For NeuralCF, the best configuration is: the number of latent factors of the GMF layer, 16; the learning rate, 0.001; the dropout rate of the MLP layers, 0.4; with the deep MLP architecture. For SSL4Rec, the best configuration is: the embedding size, 128; the dropout rate, 0.1; the learning rate, 0.001; $\alpha = 0.1$; and $\tau = 0.1$.

In the incremental learning setup, we fine-tune the incrementally updated BPR with the same set of hyper-parameters specified above except for the learning rate. The set of learning rates for fine-tuning the incremental BPR is $\{0.0025, 0.005, 0.01, 0.05, 0.1\}$. We explore smaller learning rates because the size of the training data for each incremental update is smaller, and a large learning rate with a small batch of data may result in a noisy gradient descent.

H Main Results of Experiment 1: Additional RS Algorithms

Figure H31: Impact of Data Privacy Practices with Stable User Preferences



Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

I Computational Complexity Analysis under Different Privacy Practices

We discuss the computational complexity associated with recommender systems (RS) algorithms under different privacy practices. Given that inference typically incurs much smaller computational cost than training, our discussion focuses on the training cost of RS algorithms.

To facilitate a general theoretical discussion, we abstract away the complex model architecture of neural methods and discuss three neural methods (i.e., LightGCN, NeuralCF, SSL4Rec) as a family together. We define N , M , and R as the number of users, items, and user-item interactions, respectively. Let K denote the dimensionality of the user/item latent feature space.

- **BPR:** The training complexity is $O(KR)$ (He et al. 2016), where K arises from the gradient updates on weights in the user/item latent representation and R from iterating over user-item interactions.
- **ALS:** The training complexity is $O((M + N)K^2 + RK + K^3)$ (Takács et al. 2011).
- **ItemKNN:** The training complexity is $O(M^2N)$ (Deshpande and Karypis 2004).
- **Neural Methods:** Neural methods employ the gradient descent paradigm for updating model weights during the training process. It iterates over user-item interactions to calculate the loss, and for each iteration it back-propagates to update the weights defined in the model specification. The computational complexity of this process can be generally modeled as $O(WR)$, where W is the total number of model weights. In practice, the value of W for different neural learning algorithms is a result of specific model architecture / design choices.

The data privacy practices analyzed in this study primarily affect the number of users (N) and user-item interactions (R) in the training data. They may also indirectly affect the number of items (M), if sparsely consumed items disappear from the training data due to all associated users enacting deletion. However, in this discussion, we focus on the first-order effects of privacy practices on N and R , assuming M and K remain constant.

Assume a fraction γ of users enact data deletion via the ONE_TIME or HIGH_FREQ privacy practices. Then, the post-deletion user and interaction counts become $(1 - \gamma)N$ and $(1 - \gamma)R$, respectively. As a result, model training complexity is as follows:

- BPR: $O((1 - \gamma)KR)$
- ALS: $O((M + (1 - \gamma)N)K^2 + (1 - \gamma)KR + K^3)$
- ItemKNN: $O((1 - \gamma)M^2N)$
- Neural Methods: $O((1 - \gamma)WR)$

This shows that, under deletion-based data privacy practices, increasing the fraction of affected users leads to lower model training complexity across all algorithms. The decrease of the computational complexity is essentially linear in γ across most learning algorithms, and slightly sublinear for ALS.

In contrast, under the OPT_OUT privacy practice, affected users do not delete existing data but stop adding new data in the future. In other words, as time goes by and the system accumulates more user data under NO_DELETE (but not under OPT_OUT for affected users), the computational cost of model training under NO_DELETE would be higher than under OPT_OUT. Consider a future time period, where the number of user-item interactions (which was R at the time of the privacy shock) increases by R' under NO_DELETE, i.e., a future time period with $R + R'$ interactions in total. If a fraction γ of users use OPT_OUT, then the number of total interactions that would be available to RS at that future time period would instead be $R + (1 - \gamma)R'$. Comparing the reduction of interaction data from $R + R'$ to $R + (1 - \gamma)R'$, for most learning algorithms, the model training computational complexity under OPT_OUT would reduce sublinearly in γ (as compared to the computational complexity under NO_DELETE). One exception would be ItemKNN, the asymptotic computational complexity of which would remain the same (as it depends on M and N , rather than R).

Notably, the use of data privacy practices decreases the computational cost of training RS models, due to the reduction of training data.

J Impact of Item Distinctiveness on RS Performance in the Presence of Data Privacy Practices and User Preference Changes

To further unpack the nuanced performance patterns observed under deletion-based privacy practices, particularly in scenarios where the privacy shock (T_{DataReq}) and preference shift ($T_{\text{PrefChange}}$) are temporally separated, we extend our analysis here. As noted in the main paper, HIGH_FREQ can exhibit a pronounced performance jump after a preference change—especially when $T_{\text{PrefChange}} > T_{\text{DataReq}}$. This appendix provides a deeper examination of this phenomenon by introducing and empirically validating the notion of *item distinctiveness* in users’ evolving preference structures. We show that early-stage post-change consumption is often dominated by highly distinctive items, which are easier for recommender systems to identify, even with limited recent data. This insight offers a conceptual lens to interpret the transient performance advantage of HIGH_FREQ over ONE_TIME in such settings.

We measure “distinctiveness” of item i_k at position k on user u ’s PrefList_u as:

$$D_{u,k} = \frac{p_u^{k-m} - p_u^{k+m}}{2m}$$

where $p_u^{k-m} = e_{i_{k-m}} \cdot e_u$ is user u ’s preference (i.e., relevance) score for item i_{k-m} at the $(k - m)^{\text{th}}$ position of PrefList_u . Intuitively, we measure the difference between the relevance scores of the items that are “near” the k^{th} item of PrefList_u – more specifically, between the item at m positions ahead of i_k and the item at m positions after i_k – to gauge how different (or distinguishable) the nearby items are from i_k in terms of relevance. A large difference indicates that i_k is in a neighborhood that has highly distinguishable items. Parameter m controls the size of the neighborhood in the calculation; we use $m = 2$ for this example.

Figure J32 shows the average item distinctiveness score at each rank in $PrefList_u$ across all users. On average, items that are most relevant (top 5 or 10) are indeed quite different from each other (i.e., easier for the RS to distinguish and predict), but the rest of the items (i.e., beyond the top-10 most relevant items) are much more homogeneous in terms of their relevance. Then, at each time period, we measure the average distinctiveness score of consumed recommendations²⁰ with different algorithms under HIGH_FREQ, and plot how these scores evolve in the presence of a significant user preference change. Figure J33 shows that, across all algorithms, the average distinctiveness score of consumed recommendations increases immediately after $T_{PrefChange}$, because each user is able to consume highly distinctive items among the top ranks of the user's new $PrefList_u$. This score is lower before $T_{PrefChange}$, because most highly distinctive items among the top ranks of the user's old $PrefList_u$ have already been consumed by that time.

Figure J32: Average Item Distinctiveness (Across All Users) by Relevance Rank in $PrefList_u$

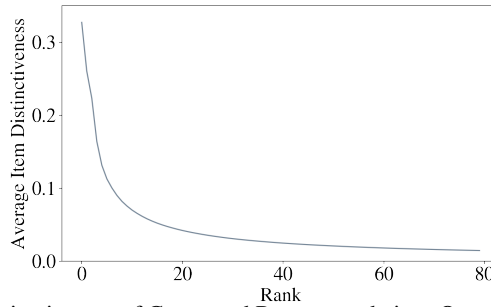
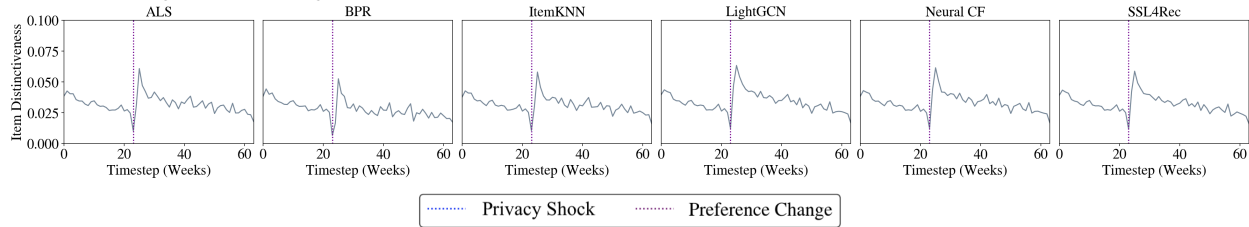


Figure J33: Average Distinctiveness of Consumed Recommendations Over Time, under HIGH_FREQ



For all graphs, the portion of the affected user population is 50%, and the degree of users' reliance on recommendations is 20%. The privacy shock happens together with the preference change. The results are for the affected user population.

²⁰Consumed recommendations represent items that were consumed by users specifically due to the RS recommendations. We obtain the average distinctiveness score by averaging the distinctiveness across all consumed recommendations in a given period.

K Main Results of Experiment 2: Additional RS Algorithms

Figure K34: Impact of Data Privacy Practices with Dynamic User Preferences

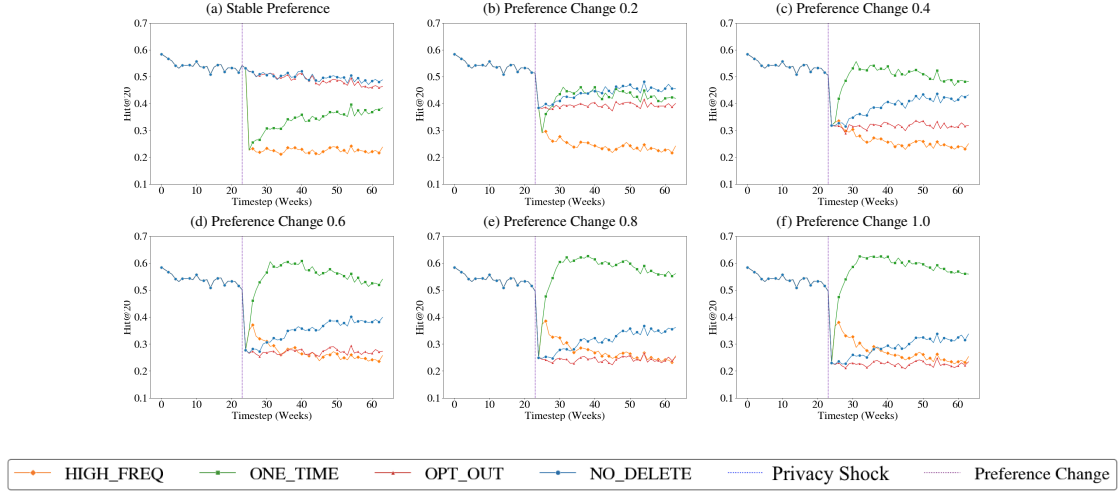


Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

L Varying the Degree of User Preference Change

This figure below shows the impact of varying the degree of user preference change (i.e., 0, 0.2, 0.4, 0.6, 0.8, 1.0) on RS performance under different privacy practices.

Figure L35: Impact of Varying Levels of User Preference Dynamic



The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

M Statistical Test Results of Post-Privacy-Shock Performance Differences

Tables M9, M10, M11, and M12 present the results of RS performance after the privacy shock. In particular, they compare the accuracy performance of the default RS setting (i.e., under NO_DELETE) to the RS performance under different data privacy practices. The RS performance is measured at 20 weeks after the privacy shock (at $T = 44$ weeks), i.e., giving some time for each configuration to reach a more stable, steady-state performance level. Each reported value represents an average across 10 simulation runs conducted with different random seeds. Statistical significance, assessed using Student's t-tests, is indicated by asterisks in the tables. These results facilitate the comparison of RS performance under various privacy practices and corroborate the visual analyses provided in the main paper.

Table M9: Performance Differences after Privacy Shock, Stable User Preferences, Hit@20

	BPR	ALS	ItemKNN	NCF	LightGCN	SSL4Rec
NO_DELETE	0.4862	0.4497	0.4015	0.4242	0.5235	0.4306
OPT_OUT	-0.0097	-0.0099	+0.0022	+0.0042	-0.0206**	0.0000
HIGH_FREQ	-0.2723**	-0.2350**	-0.1886**	-0.2092**	-0.2993**	-0.2161**
ONE_TIME	-0.1422**	-0.1349**	-0.0950**	-0.1018**	-0.1499**	-0.1172**

* and ** denotes statistical significance at 5% and 1%, respectively. RS performance is measured at $T = 44$ weeks. The NO_DELETE row shows the actual RS performance; other rows show the difference between RS performance under data privacy practices and that of NO_DELETE.

Table M10: Performance Differences after Privacy Shock, Dynamic User Preferences, Hit@20

	BPR	ALS	ItemKNN	NCF	LightGCN	SSL4Rec
NO_DELETE	0.2829	0.3835	0.2938	0.3749	0.3668	0.2604
OPT_OUT	-0.0681**	-0.1352**	-0.0519**	-0.1050**	-0.1121**	-0.0444**
HIGH_FREQ	-0.0321**	-0.1330**	-0.0550**	-0.1360**	-0.1099**	-0.0182**
ONE_TIME	+0.3194**	+0.1954**	+0.2347**	+0.1160**	+0.2006**	+0.2284**

* and ** denotes statistical significance at 5% and 1%, respectively. RS performance is measured at $T = 44$ weeks. The NO_DELETE row shows the actual RS performance; other rows show the difference between RS performance under data privacy practices and that of NO_DELETE.

Table M11: Performance Differences after Privacy Shock, IL_STANDARD, Hit@20

	$\alpha=0.0$	$\alpha=0.2$	$\alpha=1.0$
NO_DELETE	0.4862	0.4473	0.2829
HIGH_FREQ	-0.2723**	-0.2183**	-0.0321**
ONE_TIME	-0.1422**	-0.0231**	+0.3194**
IL_STANDARD	-0.0684**	-0.0657**	+0.0347**

* and ** denotes statistical significance at 5% and 1%, respectively. RS performance is measured at $T = 44$ weeks. The NO_DELETE row shows the actual RS performance; other rows show the difference between RS performance under data privacy practices and that of NO_DELETE.

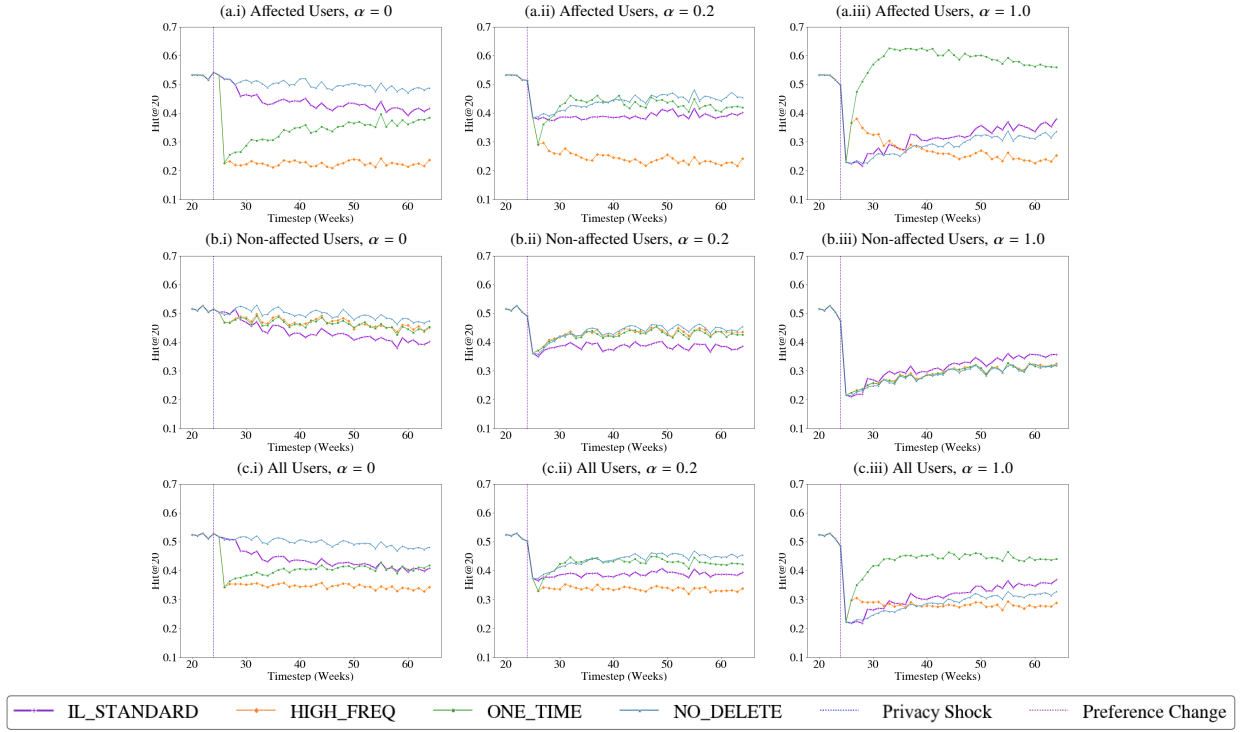
Table M12: Performance Differences after Privacy Shock, IL_OPT_OUT, Hit@20

	$\alpha=0.0$	$\alpha=0.2$	$\alpha=1.0$
NO_DELETE	0.4862	0.4473	0.2829
OPT_OUT	-0.0097	-0.0613**	-0.0681**
IL_OPT_OUT	-0.0594**	-0.0996**	-0.0714**

* and ** denotes statistical significance at 5% and 1%, respectively. RS performance is measured at $T = 44$ weeks. The NO_DELETE row shows the actual RS performance; other rows show the difference between RS performance under data privacy practices and that of NO_DELETE.

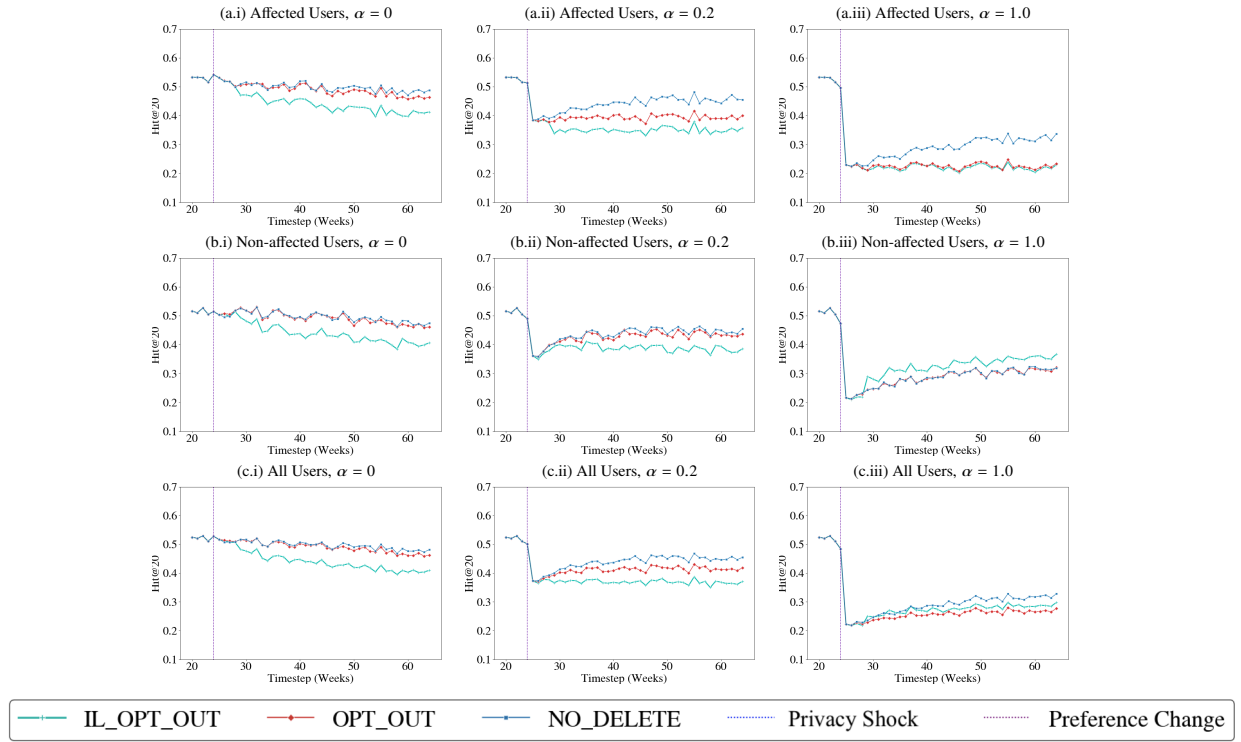
N Impact of Data Privacy Practices with Incremental Learning

Figure N36: Incremental and Non-incremental Learning under Deletion-Based Practices



The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure N37: Incremental and Non-incremental Learning under Opt-Out-Based Practices

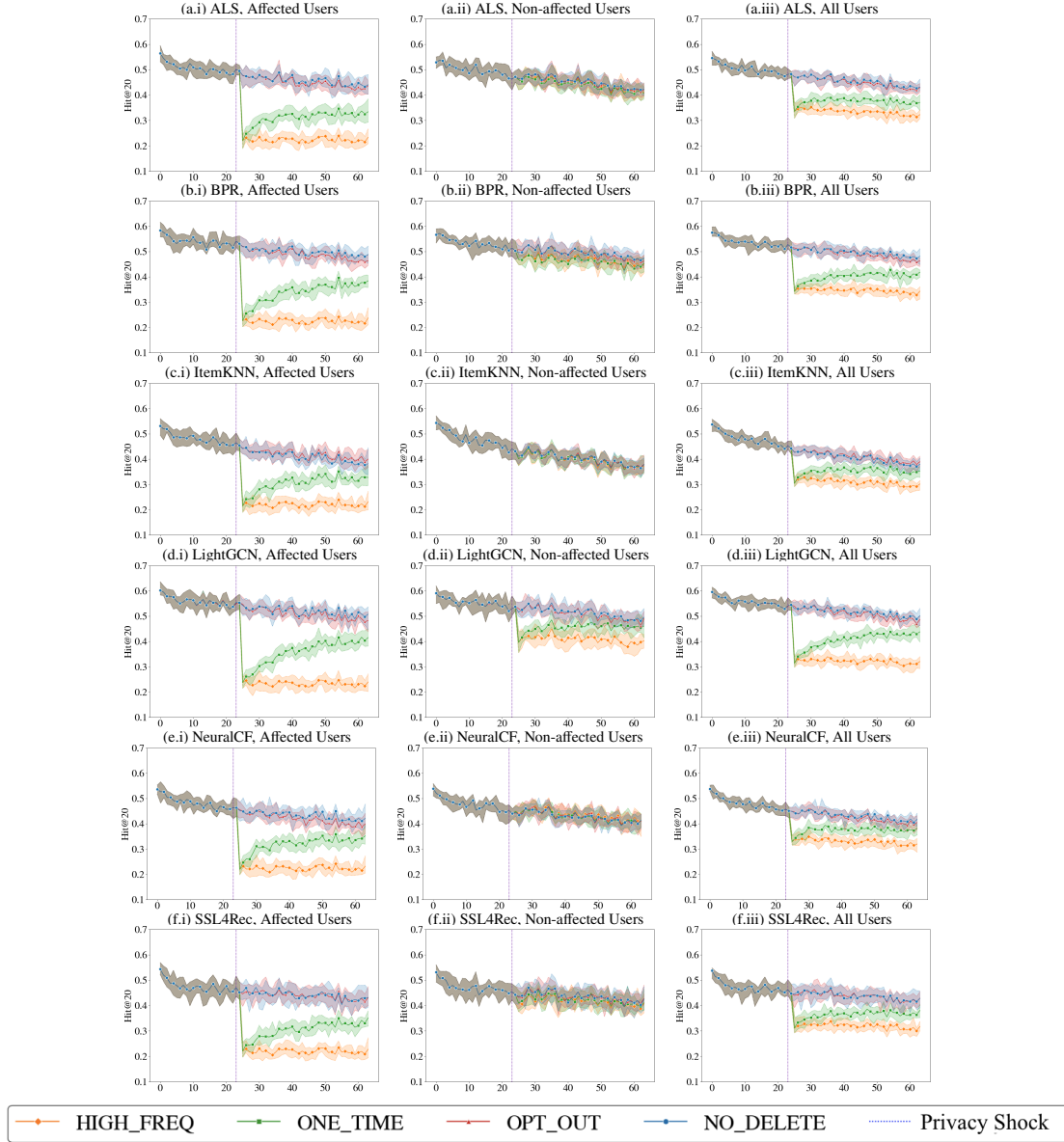


The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

O Main Experiment Results with Confidence Bands

Figures O38, O39, O40, and O41 present an alternative visualization of the main results from Experiments 1, 2, and 3, incorporating confidence bands to illustrate the degree of variability across simulation runs.²¹

Figure O38: Impact of Data Privacy Practices with Stable User Preferences, Hit@20



Note that across all figures, 50% of users are affected and all users' reliance on recommendations is 20%.

²¹The confidence bands are constructed from the maximum and minimum values of RS performance at each time step based on 10 independent simulation runs (with different random seeds).

Figure O39: Impact of Data Privacy Practices with Dynamic User Preferences, Hit@20

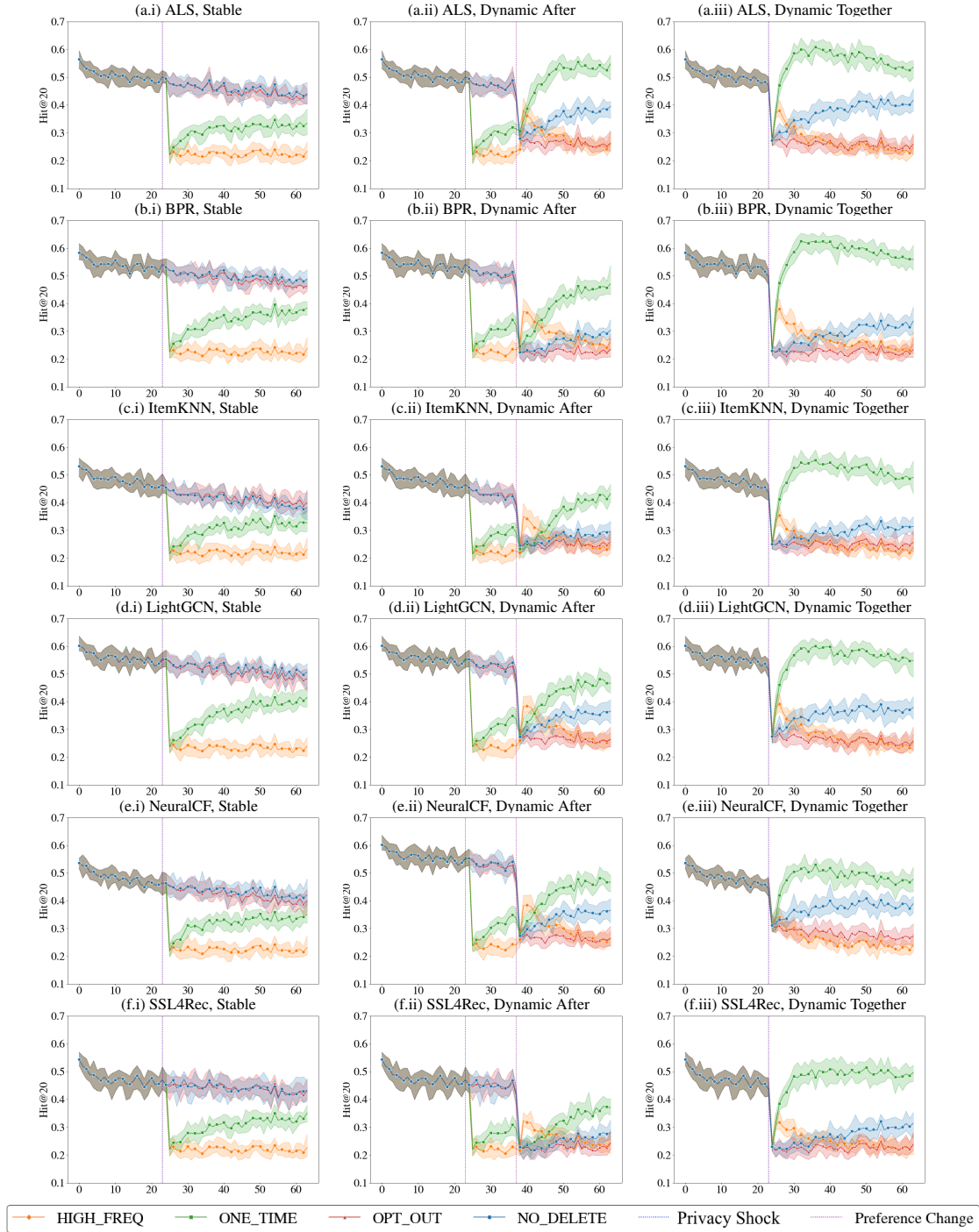
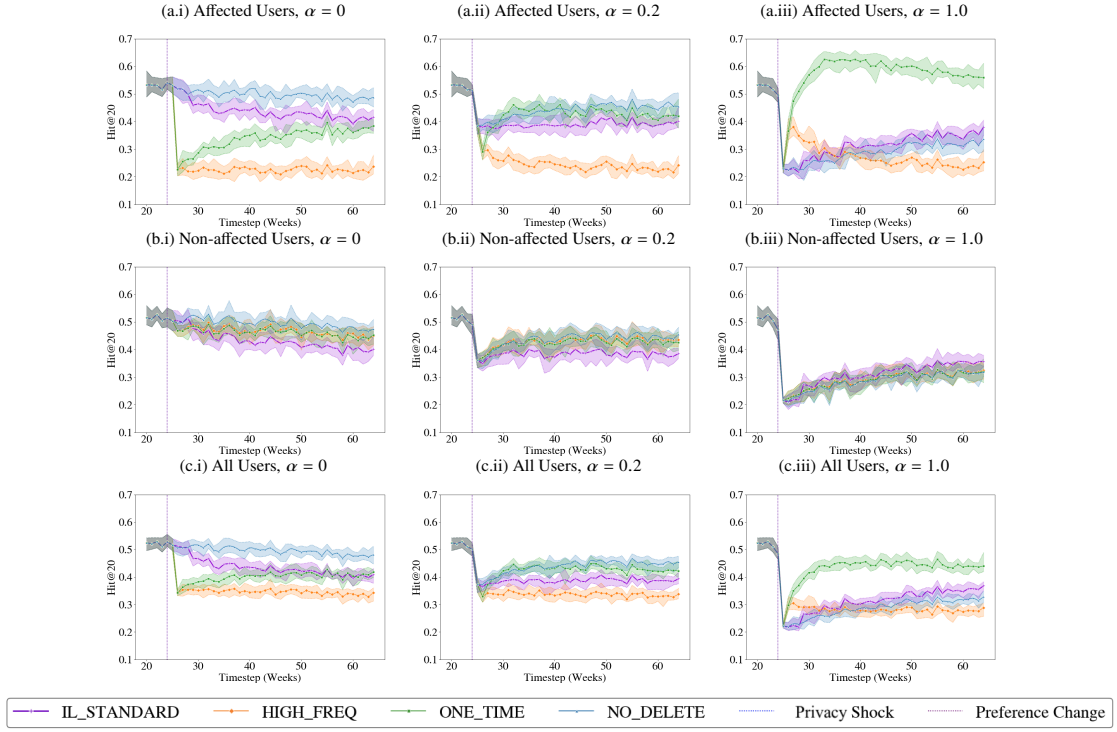
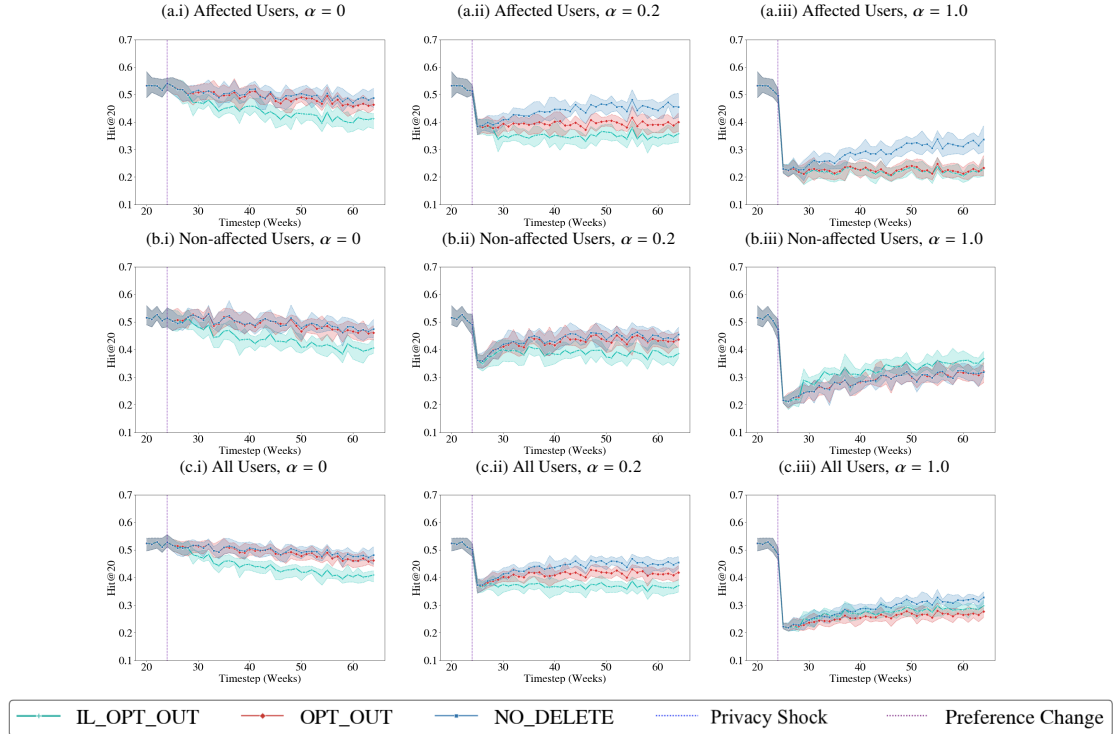


Figure O40: Incremental and Non-incremental Learning under Deletion-Based Practices, Hit@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure O41: Incremental and Non-incremental Learning under Opting Out Practice, Hit@20



The results are on affected users and the simulation is under a setting where 50% of users are affected, the algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

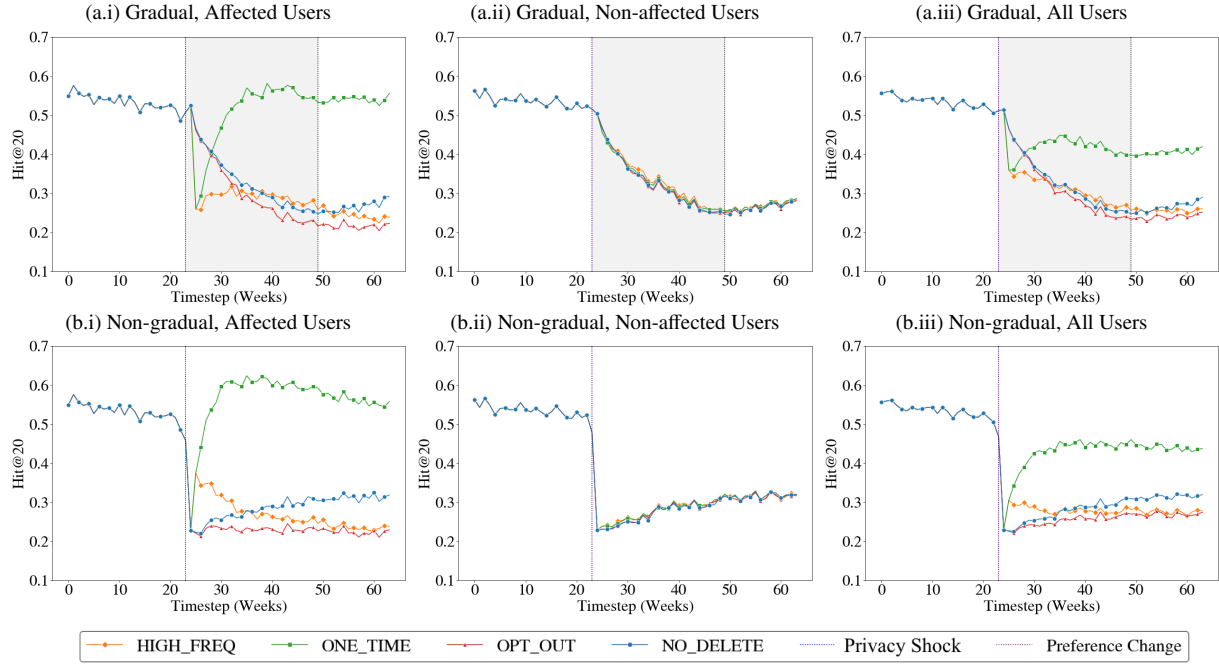
P Impact of Privacy Practice on RS Performance with Gradual Evolution of Dynamic User Preference

In Experiment 2 of the main paper, we simulated user preference change by exogenously replacing each affected user's latent preference profile with that of another randomly selected user. While this setup enables a clean and controlled investigation of how data privacy practices interact with shifts in user preferences (e.g., due to changing user's needs), it implicitly assumes that user's preference change occurs instantaneously (i.e., somewhat abruptly). However, user preferences (or needs) in real-world settings may evolve in various ways, including gradually over time (rather than shift instantaneously). To address the robustness of our findings in the settings of gradual evolution of user preferences, we use a refined simulation setup in which user preferences change incrementally following the privacy and preference change shock.

To operationalize the process of gradual preference evolution, we implemented a *progressive* adjustment of the degree of user preference change (α). Starting with $\alpha = 0$, immediately after the preference change shock we incrementally increase α by 0.04 (i.e., by 4 percentage points) each week. This incremental change continues until $\alpha = 1$ by Week 49. Once the preference change is 100% complete, α remains fixed at $\alpha = 1$ for the remainder of the simulation period. As mentioned earlier, this approach allows us to simulate a smooth and continuous shift in user preferences.

In Figure P42, we compare the results under a gradual user preference evolution setup and the original (i.e., non-gradual) user preference setup. Comparing the top and bottom rows of In Figure P42, we observe that, at the end of the simulation period, RS performance under gradual preference change with each data privacy practice converges to a level highly similar as that under non-gradual user preference setup. The results are consistent in terms of the relative rankings of RS performance for affected users under different data privacy practices, at the end of the simulation: ONE_TIME > NO_DELETE > HIGH_FREQ > OPT_OUT. This indicates that the overall, cumulative impact of data privacy practices on RS performance is highly consistent under the gradual vs. non-gradual user preference change in the medium-to-long term horizon (with some expected, logical differences during the short-term period of the gradual preference change, as discussed below).

Figure P42: Impact of Privacy Practices on RS Performance (Hit@20) with Gradual Evolution of Dynamic User Preferences



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$. In Figures (a-c), the shaded area is the period when the dynamic level gradually increases 4% weekly from 0% to 100%.

Under NO_DELETE and OPT_OUT practices, RS performance experiences a gradual decrease during the period of the gradual user preference change setting (instead of an abrupt decrease observed under the non-gradual preference change setting). This is expected, because the system retains the personal data collected before the preference change under these two privacy practices; the predictive value of this personal data gradually decreases as the user gradually shifts away from his/her old preferences. In terms of the HIGH_FREQ privacy practice, RS performance under the gradual user preference evolution does not surge after the preference change shock, although it does under the non-gradual preference change setting. As mentioned in the main paper (and discussed in Appendix J), the system performance surges because the top-ranked items in the new user preference list are more distinctive (and, hence, more accurately predictable). Now that the user preferences under the gradual evolution setting do not *immediately* switch fully to the new preferences, this distinctiveness (and, hence, the performance surge) is no longer exhibited. Finally, for ONE_TIME, we see fairly consistent patterns across the gradual vs. non-gradual preference change settings; RS performance first decreases due to the data deletion and then gradually increases as it keeps collecting more relevant (i.e., reflecting new user preferences) personal data.

Overall, as discussed above, the impact of data privacy practices on RS performance are consistent across the gradual vs. non-gradual preference change setups, which further highlights the generalizability of the key findings of the study.

Q Impact of Data Privacy Practices on RS Performance: User Population with Heterogeneous Preference Changes

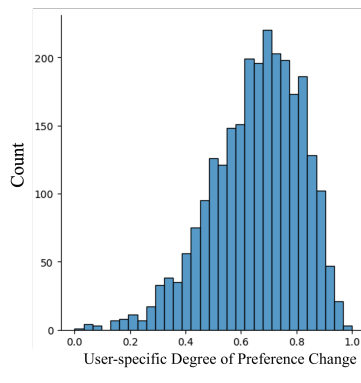
In this section, we perform an additional robustness check for our simulation setup. Specifically, our experiments in the main paper assume a *homogeneous* level of preference change in a population, i.e., all users exhibited the same degree of preference change. Here we implement a data-driven approach to derive *heterogeneous*, individual, user-specific levels of preference change that are based on the items that each user has consumed. In other words, this allows us to move from a more stylized setup, where the entire population has the same level of preference change, to a more nuanced setup, where each user exhibits their own specific level of preference change that is based on their consumption history.

To operationalize this approach, we propose a method to empirically estimate the “dynamism level” of each item based on whether (and how much) its latent representation evolves over time. To assess the item’s dynamism level, we first initialize the latent user preference profile e'_u and the latent item content profile e'_i using the first six months of our seeding dataset.²² Then, we initialize user preference profile e_u and item content profile e_i with the entire seeding dataset. We then define the dynamism level of item i as

$$ItemDynamismLevel_i = 1 - CosSim(e'_i, e_i) \quad (9)$$

where $CosSim$ is a cosine similarity measure that is normalized using the Min-Max normalization approach. If the item’s latent features change significantly over time, then the similarity between e'_i and e_i is low and, hence, the dynamism level of item i is high.

Figure Q43: Distribution of the consumption-driven, user-specific degree of preference change $\hat{\alpha}_u$

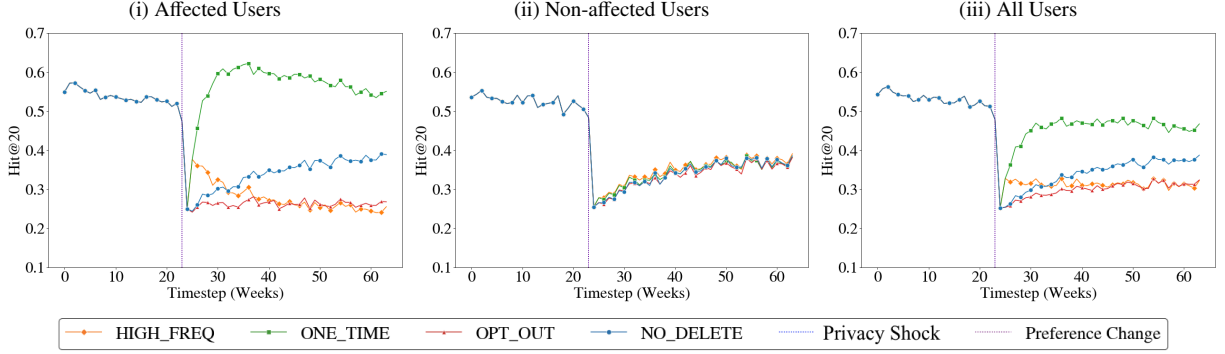


We further define the consumption-driven, user-specific degree of preference change $\hat{\alpha}_u$ of user u as the average of $ItemDynamismLevel_i$ over all items i consumed by u before the preference change. Simply put, if user u consumed more “dynamic” items, then we model the user to have a higher degree of preference change. Figure Q43 shows the distribution of the consumption-driven, user-specific degree of preference change $\hat{\alpha}_u$ across the entire population. The mean of this distribution is 0.654 with a standard deviation of

²²We use Multi-Channel Retailer seeding dataset in this exploration. This entire dataset spans the period from 2018-06-01 to 2019-07-01. We use the data in the period between 2018-06-01 to 2018-12-01.

0.161. This indicates that the user-specific degree of preference change $\hat{\alpha}_u$ is both quite high (on average) and quite diverse in the multi-channel retailer seeding dataset. This is consistent with the findings from the previous literature that user preferences can be strongly dynamic in the e-commerce context, e.g., can change significantly from visit to visit (Ludewig and Jannach 2018).

Figure Q44: RS Performance with Heterogeneous Degrees of User Preference Change

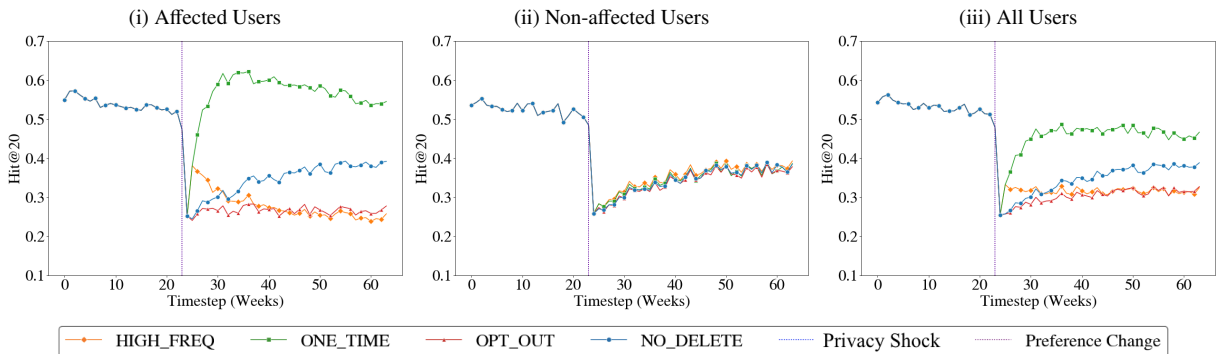


Note that across all figures, 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

In Figure Q44, we show the simulation results with the user population where each user u has an individual, consumption-driven degree of preference change $\hat{\alpha}_u$. The key RS performance patterns from Experiment 2 in the main paper, where each user had the same degree α of exogenous preference change, still hold. For visual comparison, in Figure Q45 we provide the results using the original dynamic preference setting with the homogeneous preference change degree $\alpha = 0.65$ (since the average value of $\hat{\alpha}_u = 0.654$), which shows nearly identical RS performance patterns to those in Figure Q44.

Importantly, the main findings regarding the impact of data privacy practices on RS performance remain consistent in this new setup of having heterogeneous, user-specific, consumption-driven levels of preference change, which further highlights the generalizability of the key findings of the study.

Figure Q45: RS Performance with Homogeneous Degree of Preference Change ($\alpha = 0.65$)



Note that across all figures, 50% of users are affected, algorithm is BPR, and user preferences change together with $T_{DataReq}$.

R Category-Specific Dynamic User Preferences

In Appendix Q, we consider heterogeneous, user-specific dynamic preferences where the degree of preference change depends on the item that the user consumes. In this appendix, we specifically consider one important item characteristic, the item category, which can have an impact on the degree of preference change. We extend our simulation framework by considering category-specific dynamic user preferences.

We derive the item-category-specific degree of preference change by averaging the item dynamism levels (as defined in Equation 9 of Appendix Q) across all items within the same category. Since the item dynamism level reflects how strongly user preferences on that individual item change over time, the item-category-specific degree of preference change reflects the aggregate change on items from that category. We denote the category of the item j as c_j and denote the item-category-specific degree of preference change for category c as α_c .

To model the dynamic user preference change, we first assign user u with a random new user preference profile $PrefList_v$. We then obtain ground-truth rating (i.e., preference score) in the preference profile of the user u after the preference change ($\widehat{PrefList}_u$) as

$$\hat{p}_{uj} = (1 - \alpha_{c_j}) \times p_{uj} + \alpha_{c_j} \times p_{vj} \quad (10)$$

where p_{uj} is the preference score for item j in user u 's original preference profile ($PrefList_u$), and p_{vj} is the preference score for item j in user v 's preference profile ($PrefList_v$). This equation characterizes the intuition behind the item-category-dependent preference change: when an item belongs to a category with a stronger dynamism level (α_{c_j} is higher), the new preference score ($\hat{p}_{uj}$) deviates more strongly from the original preference score (p_{uj}). In essence, this model captures the idea that user preferences can be more dynamic for certain item categories than others.

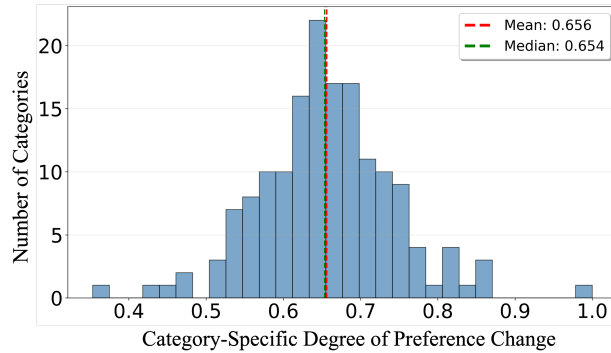
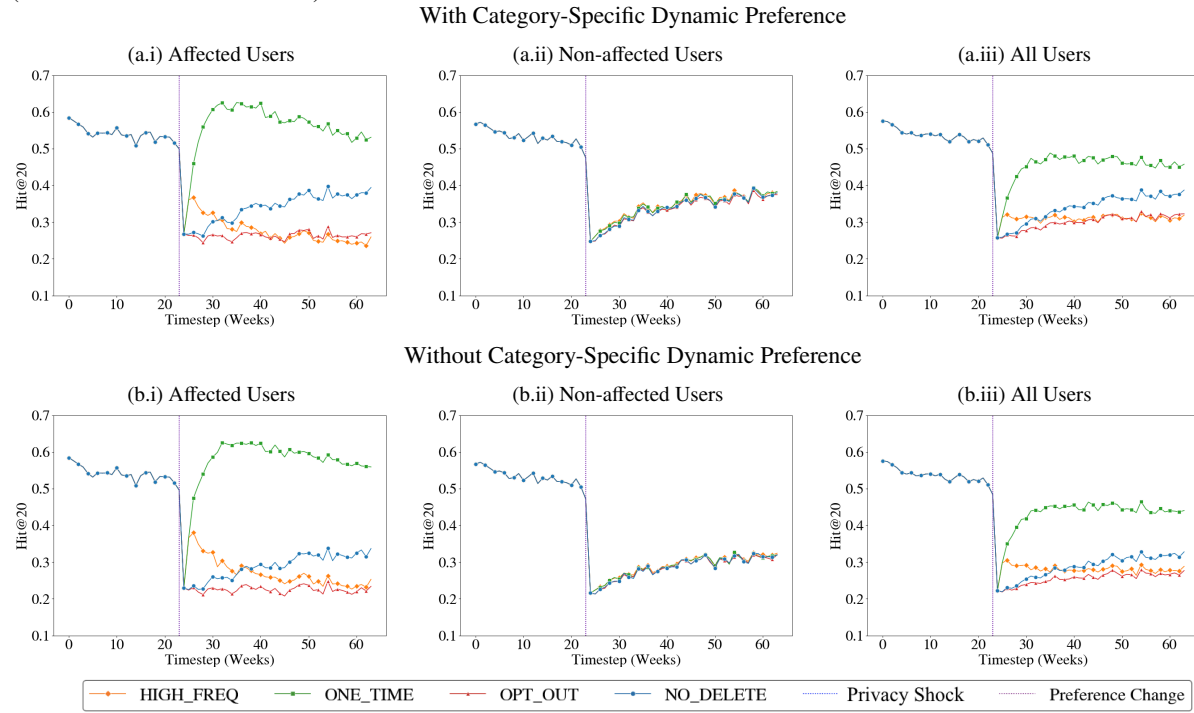


Figure R46: Histogram of Category-Specific Degree of Preference Change

We obtain the item category information for the Multi-Channel Retailer dataset. There are 159 categories in total across all items in the simulation. We visualize the distribution of the category-specific degree of preference change in Figure R46. The category-specific degree of preference change is 0.65 on average. We then simulate the main experiment setup for dynamic user preferences (i.e., in Experiment 2 of the main paper) with this category-specific dynamic preference modeling. Figure R47 presents a comparison on simulation results between the setup with category-specific dynamic preference modeling and the standard dynamic

Figure R47: Impact of Privacy Practices on RS Performance (Hit@20) with vs. without Category-Specific Dynamic Preference (Multi-channel Retailer Dataset)



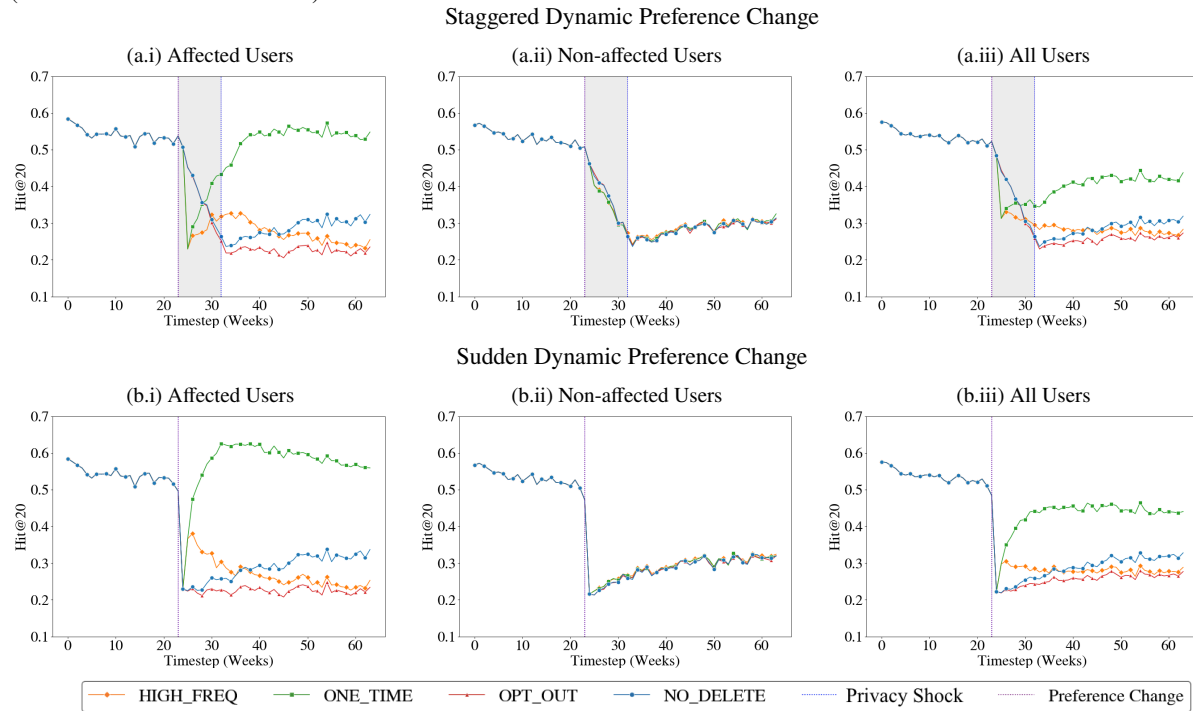
The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$. The simulation without category-specific dynamic preference sets a homogeneous degree of preference change at 100% for all users.

preference setup (as in Experiment 2 of the main paper). We observe that the general performance patterns are highly consistent across two setups, demonstrating strong robustness and generalizability of main findings in the paper.

S Staggered User Preference Change

We relax the assumption that the user preference change is a one-time, sudden shock by simulating a staggered user preference change. We randomize the users into ten groups, corresponding to the first ten weeks after the deletion shock. The user preferences for each group change during the assigned week. This setup maintains an experimentally controlled preference change while mimicking the non-synchronicity of the user preference change that can be more representative of the real-world settings. Figure S48 compares the simulation results for the staggered vs. sudden preference change settings. As would be expected, the performance degradation under NO_DELETE and OPT_OUT and the performance recovery under ONE_TIME become comparatively more gradual during the period of the staggered preference change. In terms of the system performance under HIGH_FREQ, the sudden surge after the privacy shock also becomes more gradual and smooth. After the staggered change period, the main patterns are highly consistent with those in the sudden preference change setup (used in the main paper).

Figure S48: Impact of Privacy Practices on RS Performance (Hit@20): Staggered vs. Sudden Dynamic Preference Change (Multi-channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$. The simulation sets a homogeneous degree of preference change at 100% for all users. The gray area highlights the period of the staggered dynamic preference change.

T Calibrating User Reliance on Recommendations

We calibrate user reliance on recommendations based on real-world data. Specifically, we use the click-stream data of users in the Multi-Channel Retailer seeding dataset. The click-stream data records different types of user actions, such as clicks on the recommendation list and clicks on product pages. We estimate the user reliance on RS as the percentage of clicks on the recommendation list for each user. Figure T49 shows the distribution of the empirical user reliance on RS. There were 851 users who never clicked on the recommendation list. The average percentage of clicks on the recommendation list (across all users) is 0.145 with a standard deviation of 0.19.

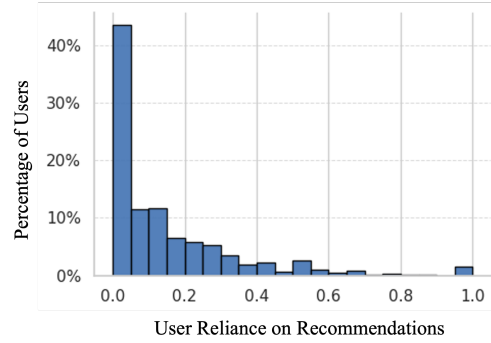
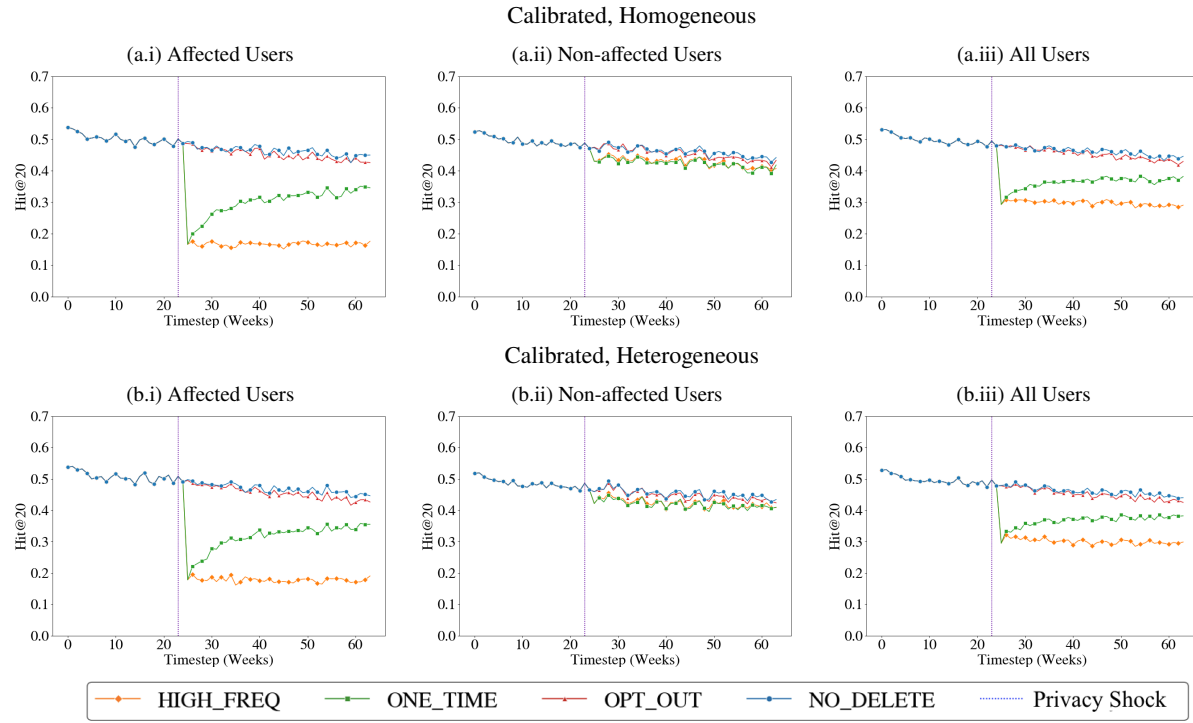


Figure T49: Empirical User Reliance on RS

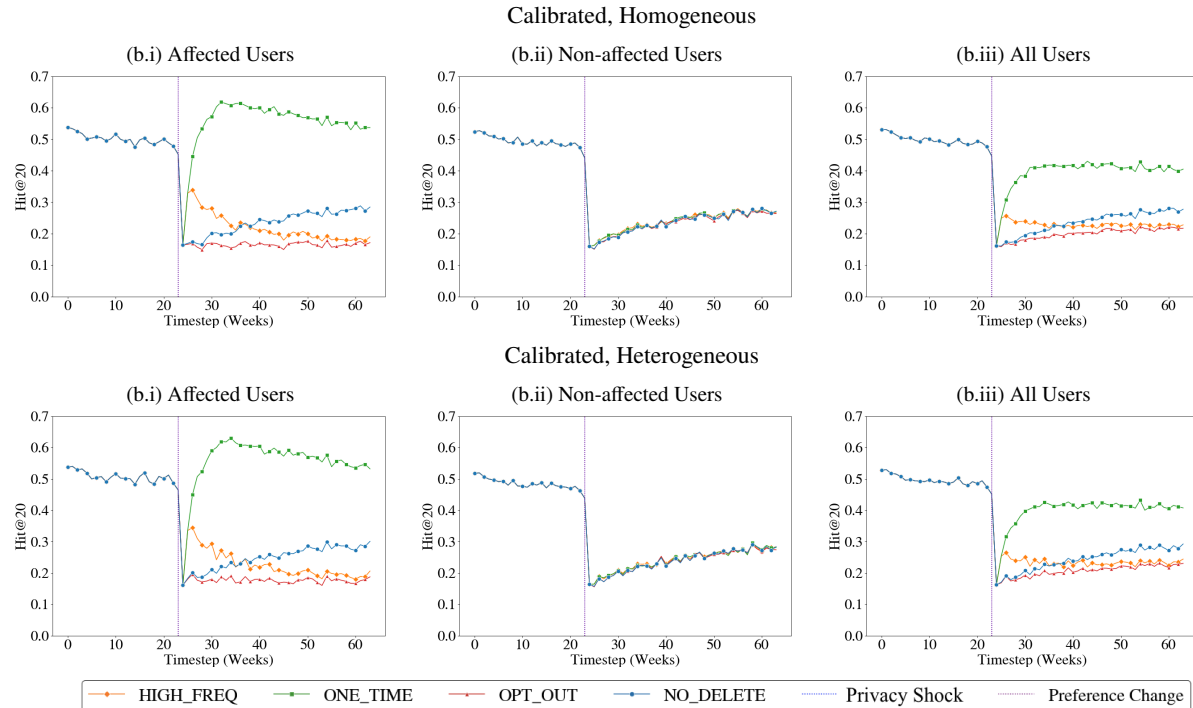
We run two versions of simulations calibrated to the empirical user reliance on recommendations. Specifically, in the first “Calibrated, Homogeneous” version, we calibrate the global, fixed user reliance to the average of the empirical user reliance (i.e., 14.5%); in the second “Calibrated, Heterogeneous” version, we calibrate each user’s reliance to its corresponding empirically estimated value (in a heterogeneous manner). Figures T50 and T51 show comparisons between simulation results of Calibrated Homogeneous vs. Heterogeneous reliance settings under stable and dynamic user preferences, respectively. The general RS performance patterns in both versions are highly consistent to those observed in the main paper, demonstrating strong robustness and generalizability of the main findings with realistic considerations.

Figure T50: Impact of Privacy Practices on RS Performance (Hit@20) with Calibrated Homogeneous vs. Heterogeneous User Reliance on Recommendations (Stable User Preferences, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure T51: Impact of Privacy Practices on RS Performance (Hit@20) with Calibrated, Homogeneous vs. Heterogeneous User Reliance on Recommendations (Dynamic User Preferences, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

U Dynamic User Reliance on Recommendations

We relax the assumption that user reliance on recommendations is static and homogeneous across users by introducing a dynamic and user-specific reliance model. We model the probability of a user relying on RS, r_t , using a time-dependent logistic function. Let $\tau^u(t)$ denote the most recent timestep before t at which user u consumed a recommended item, and let $p_{u, \tau^u(t)}^{rec}$ be the true preference score of that item. The user-specific dynamic reliance is defined as:

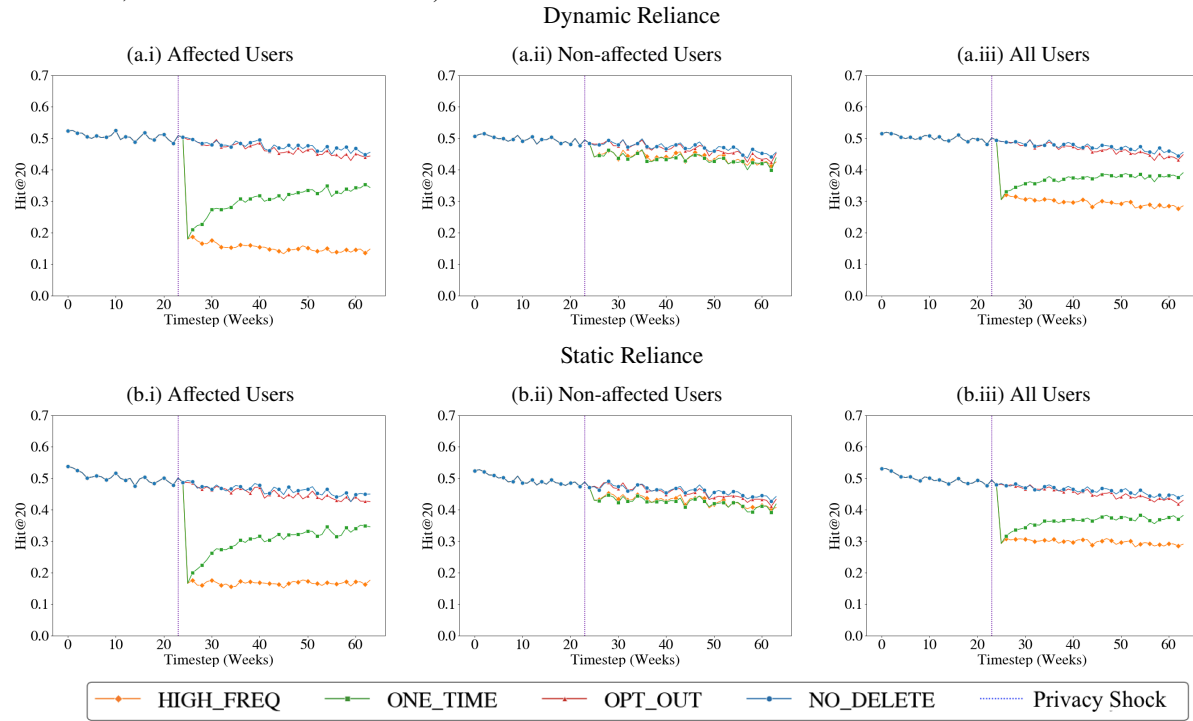
$$r_t^u = \sigma\left(\theta_0 + \theta_1 p_{u, \tau^u(t)}^{rec}\right) \quad (11)$$

where intercept θ_0 governs the base reliance probability. Reliance r_t^u is piecewise constant between consecutive recommendation-consumption events. In other words, the degree of user reliance on recommendations can only change when the user consumes from the recommended items list; this reliance remains unchanged if the user does not consume or consumes items not based on recommendations. In the case where the user consumes one of the recommended items, if the consumed item has low relevance for the user (i.e., the user’s ground-truth rating is low), the user reduces their reliance on RS. Thus, coefficient θ_1 can be interpreted as the user’s sensitivity to the system performance change, and coefficient θ_0 is the baseline reliance due to the user’s algorithmic appreciation or aversion (Logg et al. 2019, Dietvorst et al. 2015). This perspective that user reliance can be dynamically adjustable based on user’s experience of system performance is supported in previous studies on user trust in machine learning systems (Lohrenz et al. 2007, Yu et al. 2017, 2018).

We simulate the main experiment setups (i.e., the stable vs. dynamic user preference setups of Experiments 1 and 2 in the main paper) by replacing the static user reliance with dynamic user reliance model for the Multi-Channel Retailer seeding dataset. We set $\theta_0 = -2$ and $\theta_1 = 0.9$ to calibrate the user reliance (before the privacy shock) based on the one observable in the real world (14.5% based on our estimation for Multi-channel Retailer Dataset; please see Appendix T for details). We show simulation results under this setup of dynamic user reliance in Figures U52 and U53.

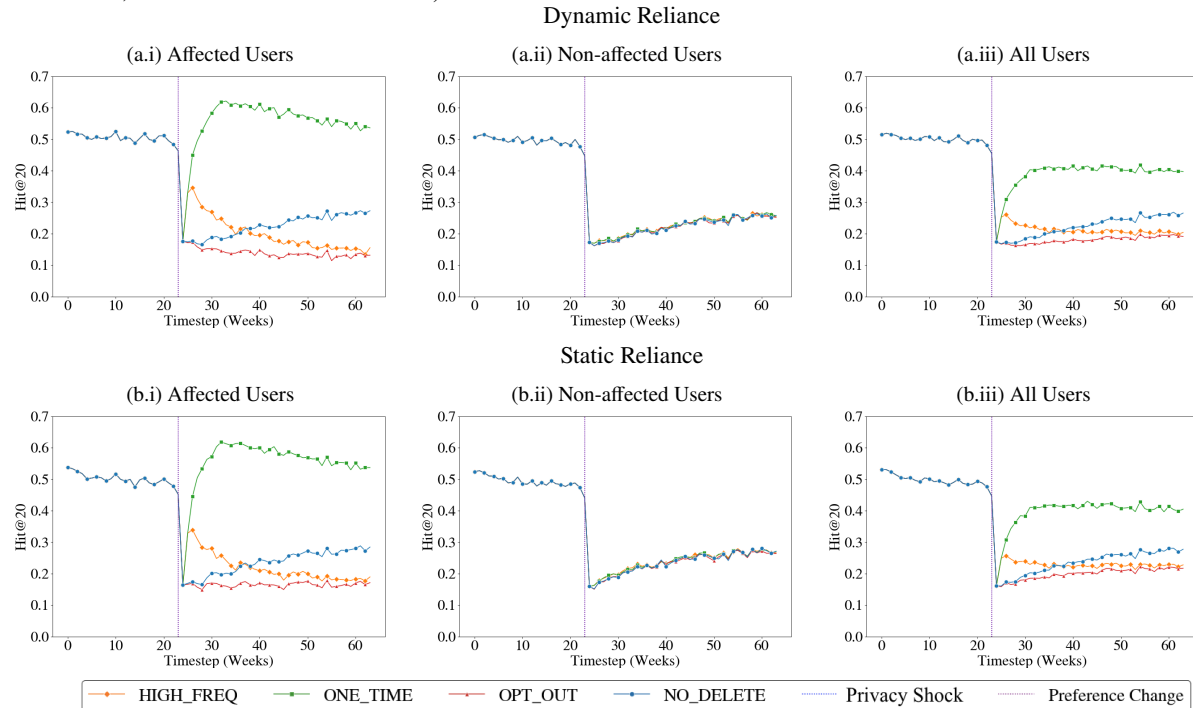
Figures U52 and U53 compare simulation results for dynamic vs. static RS reliance settings under stable and dynamic user preferences, respectively. The general performance patterns remain highly consistent across the the dynamic vs. static RS reliance settings under both stable and dynamic user preferences. We also plot the average dynamic user reliance over time in Figure U54. We observe that the user reliance increases gradually before the privacy shock because the system is able to recommend more relevant items with the accumulation of more interaction (training) data. Then, the user reliance drops (by different degrees for different privacy practices) after the privacy shock because the system performance degrades. This observation provides further support that the dynamic user reliance modeling properly reflects the changes in performance-based user trust in the system.

Figure U52: Impact of Privacy Practices on RS Performance (Hit@20) with Dynamic vs. Static User Reliance (Static User Preferences, Multi-Channel Retailer Dataset)



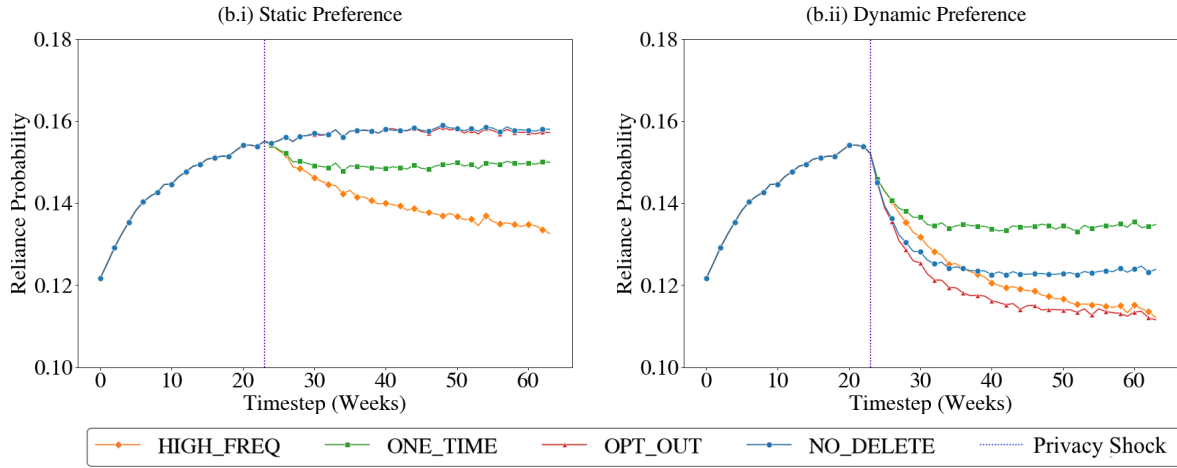
The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure U53: Impact of Privacy Practices on RS Performance (Hit@20) with Dynamic vs. Static User Reliance (Dynamic User Preferences, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure U54: Dynamic User Reliance by Timesteps (Static vs. Dynamic User Preferences, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

V Modeling New Item Arrival

We extend our simulation framework to incorporate the arrival of new items. To accurately reflect the real-world item dynamics, we seeded the simulated item arrival schedule based on the arrival patterns drawn from the original dataset. Since the dataset does not directly record item creation dates, we used the item's *first-seen* date (the date it was first consumed by any user) as a practical proxy for its arrival on the platform.

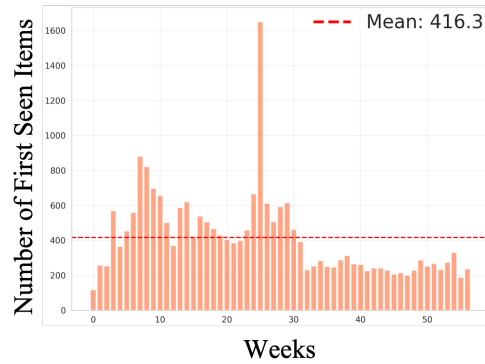


Figure V55: Number of First-Seen Items by Week

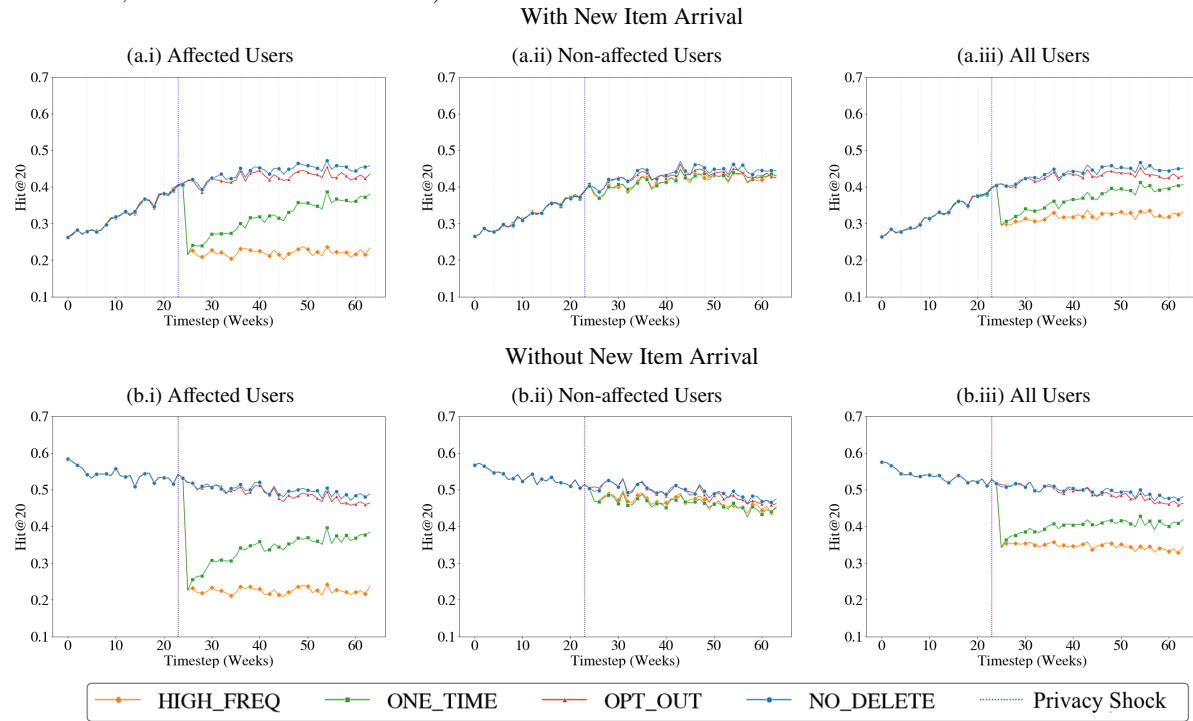
Figure V55 shows the number of newly arrived items in each week for the Multi-Channel Retailer seeding dataset. On average, around 416 items arrive each week.²³ As mentioned earlier, we consider the first-seen date of an item as its arrival date. However, some items may have actually arrived before the period covered by our dataset; therefore, the number of arrivals during the early portions of our data period could be overestimated because of that. For this reason, we perform the arrival rate estimation only based on the later portion of our data period (i.e., after 32nd week). Figure V55 shows that the number of first-seen items

²³The dataset reflects the period from June 1, 2018 to July 1, 2019. The 25th week in this period was the week of Black Friday, when the sales volume increased dramatically and the number of first-seen items also spiked strongly.

in each week is almost uniformly distributed around the mean of 205, which is roughly 0.8% of items in the entire item population of the seeding dataset. Thus, in our simulation, we also adopt a uniform item arrival schedule with 0.8% of the entire item population arriving each week (i.e., 20 new items arrive uniformly each week for 65 weeks). To avoid the strict cold-start problem (i.e., where there is zero interaction information about the item), these new items arrive in the system with 5 consumption records.²⁴

In Figures V56 and V57, we compare simulation results for the settings with vs. without new item arrival under stable and dynamic user preferences, respectively. The general performance patterns are highly consistent with one nuance. We observe that the baseline system performance with new item arrival has a general tendency to increase over time, because new item arrival continuously reveals richer user preference information, which the system learns to improve its performance gradually. In contrast, the baseline system performance without new item arrival has a general tendency to slightly decrease over time. This reflects the fact that, over time, the most-preferred items get consumed by users and, without new item arrival, the quality of the remaining items will inevitably slowly decrease.

Figure V56: Impact of Privacy Practices on RS Performance (Hit@20) with vs. without New Item Arrival Modeling (Stable User Preferences, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

²⁴This is consistent with the real-world practice where strict cold-start items are excluded from RS until they have accumulated some minimal number of interactions that can then be used for personalization purposes.

Figure V57: Impact of Privacy Practices on RS Performance (Hit@20) with vs. without New Item Arrival Modeling (Dynamic User Preferences, Multi-Channel Retailer Dataset)



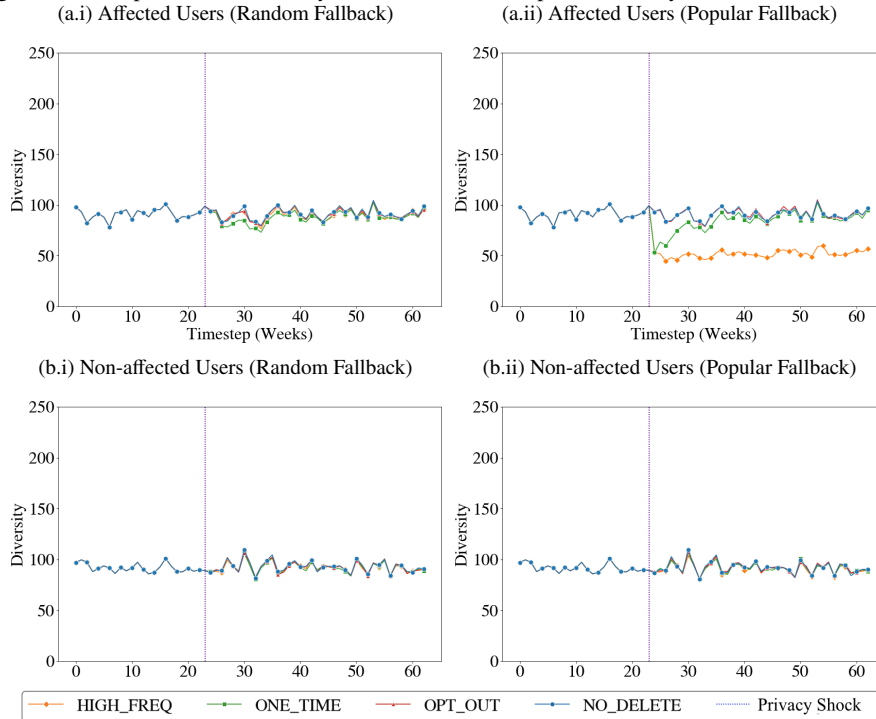
The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

W Impact of Data Privacy Practices on Consumption Diversity

Although the main focus of our study is the impact of data privacy practices on the RS accuracy performance, data privacy practices may influence various item consumption patterns as well. Therefore, in this section we explore the impact of data privacy practices on the diversity of consumed items.

In particular, we focus on a popular measure of *aggregate diversity*, which is often measured by the total number of distinct consumed items (Adomavicius and Kwon 2011, Zhang et al. 2020), and we use a period of one week as a temporal granularity at which we measure consumption diversity over time.²⁵ As part of our explorations under the stable user preference setting, as shown in Figure W58, we find that the consumption diversity does not change significantly under OPT_OUT. However, the impact of deletion-based practices (ONE_TIME and HIGH_FREQ) can vary substantially, primarily based on what *fallback* methods are used.²⁶ We illustrate this using two fallback methods, *random* fallback and *popular* fallback, in Figure W58. Specifically, to users who do not have sufficient historical consumption data (due to their enacted privacy actions), random fallback recommends random items, while popular fallback recommends the top-20 items that are currently most popular across the entire user population (based on the historical consumption data that is available to RS). For users who have sufficient consumption data, we use the intended RS algorithm to provide personalized recommendations.

Figure W58: Impact of Data Privacy Practices on Consumption Diversity, Stable User Preferences



Note that across all figures, the algorithm is BPR, 50% of users are affected, and users' reliance on recommendations is 20%.

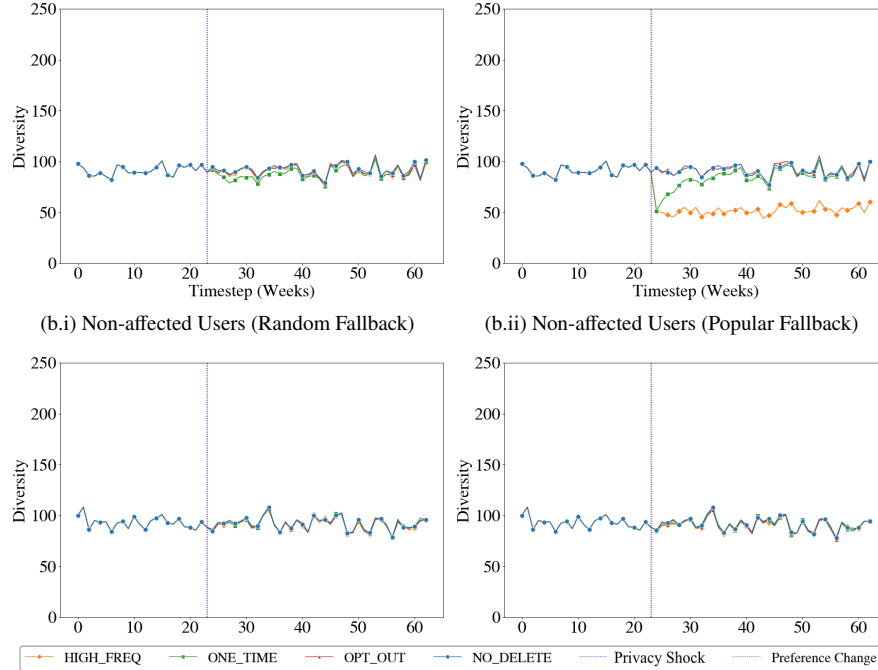
²⁵To isolate the effect on consumption diversity solely due to RS, we focus only on the items that were consumed due to recommendations; other (e.g., need-based) consumptions are not influenced by RS.

²⁶A fallback method is commonly used when the primary recommendation algorithm cannot produce meaningful results, typically because there is insufficient data about a user or item.

Results in Figure W58(a) show that, with random fallback, consumption diversity under HIGH_FREQ and ONE_TIME does not change substantially after the privacy shock; however, consumption diversity decreases significantly with popular fallback, as shown in Figure W58(b). This decrease is attributed to more concentrated recommendations on popular items due to the use of popular fallback, leading to more focused consumption on a smaller subset of items. Following the decrease, consumption diversity under ONE_TIME gradually recovers, as the use of popular fallback is reduced due to the alleviated data sparsity with gradual accumulation of personal data. In contrast, random fallback maintains the diversity near the maximum possible level throughout the simulation period due to completely random recommendations.

We also note that consumption diversity for non-affected users does not change substantially throughout the simulation period, as shown in W58(c-d), indicating a minimal spillover effect. Moreover, consumption diversity patterns are highly consistent across the stable and dynamic user preference settings, as can be seen by comparing Figures W58 and W59.

Figure W59: Impact of Data Privacy Practices on Consumption Diversity, Dynamic User Preferences
(a.i) Affected Users (Random Fallback) (a.ii) Affected Users (Popular Fallback)



Note that across all figures, the algorithm is BPR, 50% of users are affected, and users' reliance on recommendations is 20%.

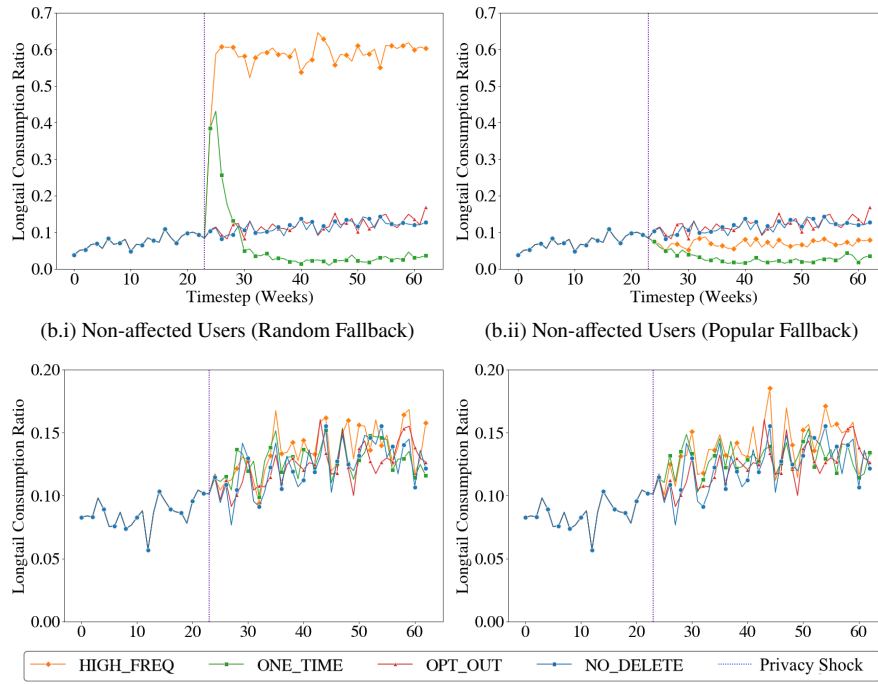
X Impact of Data Privacy Practices on Long-Tail Item Consumption

We define the set of long-tail items $I_{longtail}^t$ as the 80% least popular items in the training data of the RS at time t .²⁷ We measure the long-tail consumption ratio (LCR) as follows:

$$LCR_t = \frac{\sum_{u \in U} |TrainData_u^t \cap I_{longtail}^{t-1}|}{\sum_{u \in U} |TrainData_u^t|}.$$

Note that item consumptions at t (i.e., $TrainData_u^t$) are made under the potential influence of RS, which was trained on data collected up to (and including at) $t - 1$. Thus, in the above definition, we categorize items as “long-tail” based on the training data collected at $t - 1$ to reflect what RS perceives as long-tail.

Figure X60: Impact of Data Privacy Practices on Long-Tail Item Consumption, Static User Preferences
(a.i) Affected Users (Random Fallback) (a.ii) Affected Users (Popular Fallback)



Note that across all figures, the algorithm is ALS, 50% of users are affected, and users' reliance on recommendations is 20%.

Figure X60 shows the long-tail consumption ratio under the stable user preference setting with two different fallback methods: random and popular. We find that the consumption longtailness does not change significantly under OPT_OUT, which remains consistent with the performance patterns measured by recommendation accuracy and consumption diversity. However, the impact of deletion-based practices (ONE_TIME and HIGH_FREQ) varies depending on the fallback method that is used by a given RS algorithm. With random fallback, long-tail item consumption of affected users under HIGH_FREQ and ONE_TIME rises right after the privacy shock. It is because RS loses personal data of affected users, and random fallback by

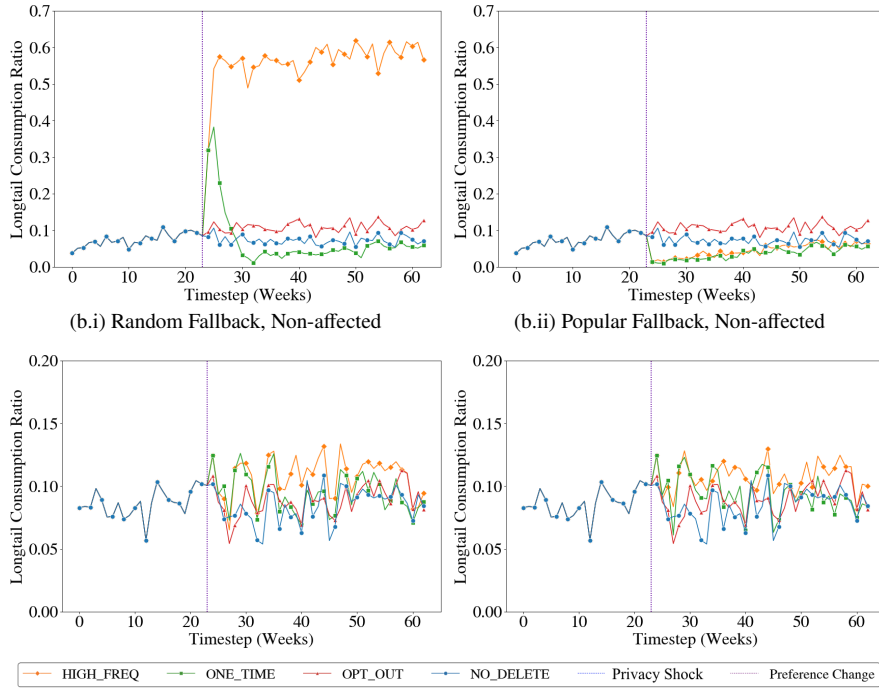
²⁷We also explored using 70% and 90%, with results remaining qualitatively consistent.

definition will recommend items that have 80% probability of being long-tail. The long-tail item consumption ratio under ONE_TIME in Figure X60(a) gradually decreases after the initial surge as the personal data accumulates, due to the fact that many RS algorithms tend to recommend more popular items when having smaller amounts of data for personalized inference. In Figure X60(b), the long-tail item consumption drops significantly under ONE_TIME and HIGH_FREQ immediately after the privacy shock because popular fallback recommends popular items.

We also note that long-tail item consumption by non-affected users is not significantly impacted by data privacy practices, as shown in Figures X60b.i and X60b.ii, indicating a minimal spillover effect.

Long-tail item consumption under the dynamic user preference setting share highly similar patterns, as shown in Figure X61.

Figure X61: Impact of Data Privacy Practices on Long-Tail Item Consumption, Dynamic User Preferences
(a.i) Random Fallback, Affected (a.ii) Popular Fallback, Affected



Note that across all figures, the algorithm is ALS, 50% of users are affected, and users' reliance on recommendations is 20%.

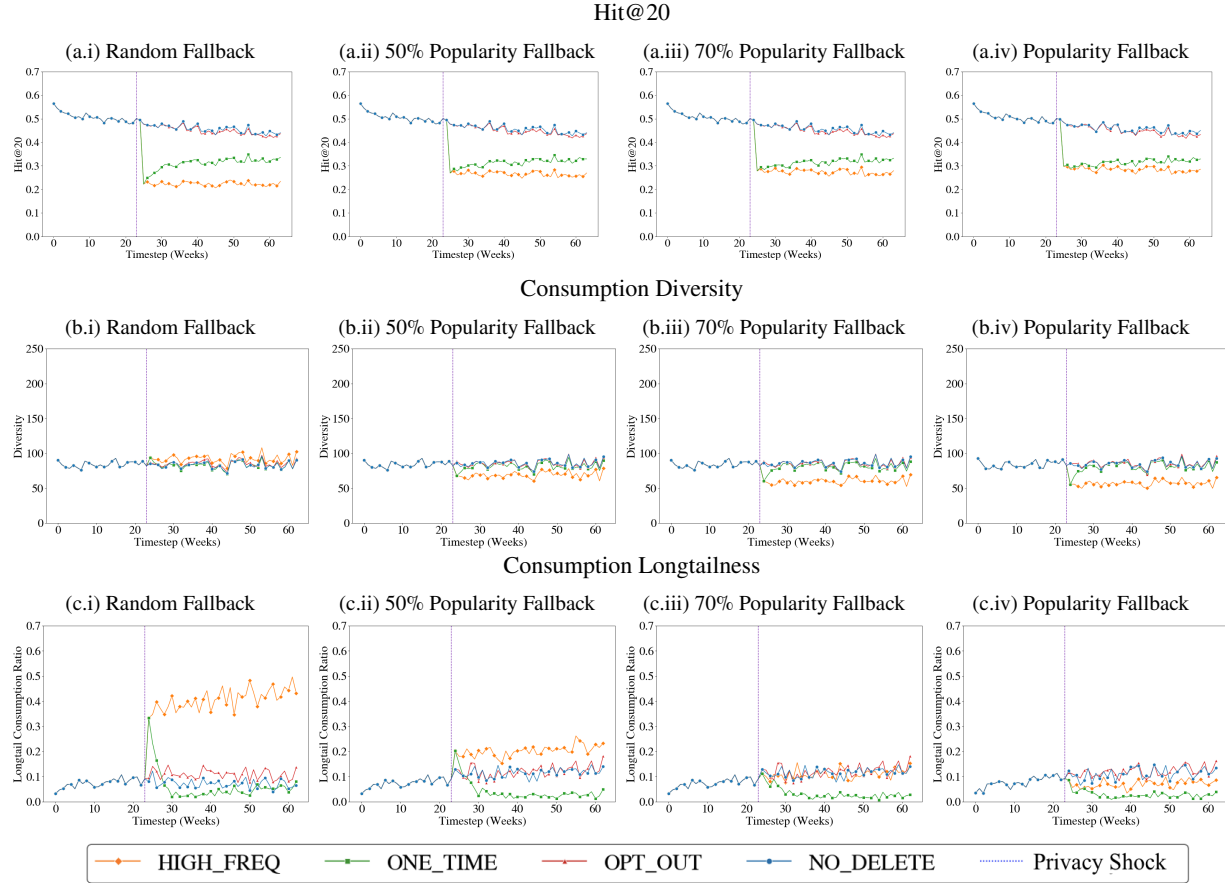
Y Hybrid Fallback Methods

Appendices W and X show that the consumption diversity and longtailness can be strongly affected by the fallback methods, such as random fallback or popularity fallback. In this appendix, we further explore hybrid fallback strategies and assess the impact of different fallback strategies on key metrics of system performance (i.e., Hit@20) and consumption dynamics (i.e., diversity and longtailness).

In particular, we examine 4 types of fallback methods for users who do not have sufficient personal data for personalized recommendations: the Random Fallback (i.e., 100% of recommendations are randomly selected items), 50% Popularity Fallback (i.e., 50% of recommendations are popular items and 50% random items),

70% Popularity Fallback (i.e., 70% popular and 30% random), and Popularity Fallback (i.e., 100% popular). We run simulations with the Multi-Channel Retailer seeding dataset and report results in Figures Y62 and Y63.

Figure Y62: Impact of Privacy Practices on RS Performance (Hit@20) with Different Fallback Methods (Stable User Preferences, Affected Users, Multi-Channel Retailer Dataset)



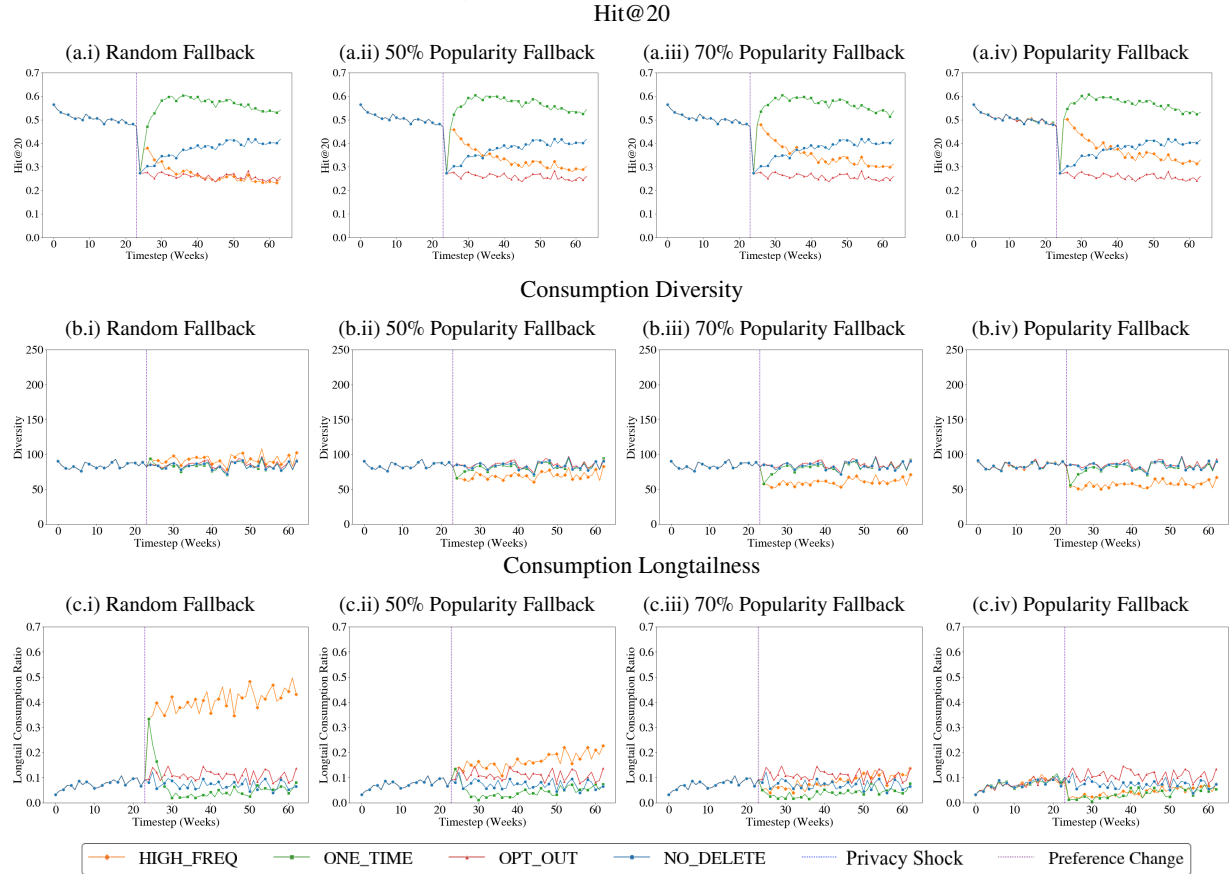
The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Figure Y62 shows the simulation results under stable user preferences for different fallback methods. We observe that the impact of fallback methods mainly differs for the HIGH_FREQ practice, under which most affected users do not have any personal data, and thus the system needs to rely on fallback methods for recommendations. The choice of fallback methods hardly has any impact on system performance and consumption dynamics under other privacy practices because the system rarely rely on fallback methods in those scenarios. Importantly, moving from column 1 to column 4, we observe that the system performance under HIGH_FREQ improves as fallback methods include increasingly larger portion of popular items. In contrast, the consumption diversity and longtailness of HIGH_FREQ decrease as more popular items are recommended. This is consistent with previous observations in Appendices W and X, confirming that popularity fallback methods tend to improve the system performance at the cost of consumption diversity and longtailness.

Figure Y63 shows the simulation results under dynamic user preferences for different fallback methods. The aforementioned HIGH_FREQ performance patterns are highly consistent across the stable and dynamic

preference settings, validating robustness of the fallback-method-related findings.

Figure Y63: Impact of Privacy Practices on RS Performance (Hit@20) with Different Fallback Methods (Dynamic User Preferences, Affected Users, Multi-Channel Retailer Dataset)



The results are under a setting where 50% of users are affected, algorithm is BPR, and the user preference changes together with $T_{DataReq}$.

Z External Validity of Simulated User Consumption Behavior

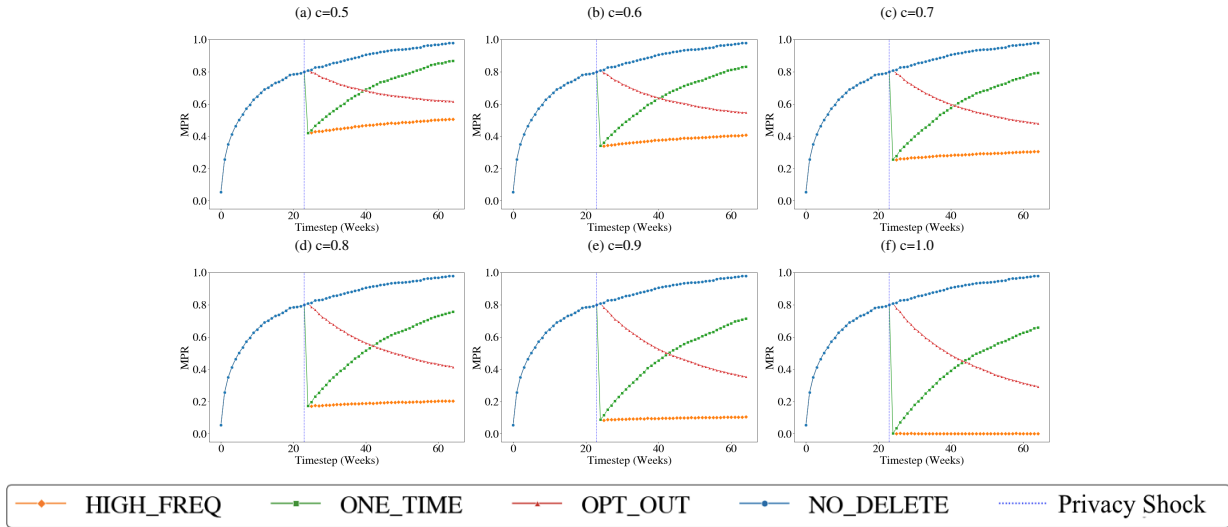
In this appendix, we show that the simulated user consumption behavior is a faithful reflection of real-world consumption patterns. We examine two fundamental aspects of consumption behavior: users' global ranking by their consumption frequency and items' global ranking by popularity. Together, these aspects provide an additional assessment of the external validity of the simulated user consumption behavior.

We compare the aforementioned consumption aspects in (a) the real-world seeding dataset vs. (b) the simulated dataset at time 0, before the intervention of a recommender system (i.e., with simulated dataset $ConsData_u^0$, the simulated consumption records without recommendations). Specifically, we obtained the users' global ranking by consumption frequency on these two datasets and assessed the rank correlation between them; same analysis is performed for items' global ranking by popularity. The ranking correlation (Spearman's r) of users' global ranking by consumption frequency is 0.8968 and that of items' global ranking by popularity is 0.8084, both statistically significant with $p\text{-value} < 0.001$. This shows that the simulated user consumption behavior strongly preserves the relative consumption frequency and item popularity from the seeding dataset on a global scale. These statistics together demonstrate that the simulated consumption behavior comprehensively reflects the real-world consumption behavior (observed in the seeding dataset), building strong external validity into the consumption dynamics of the simulation environment.

AA Parameterizable Privacy Practices: Impact of Niche Threshold on the MPR Metric

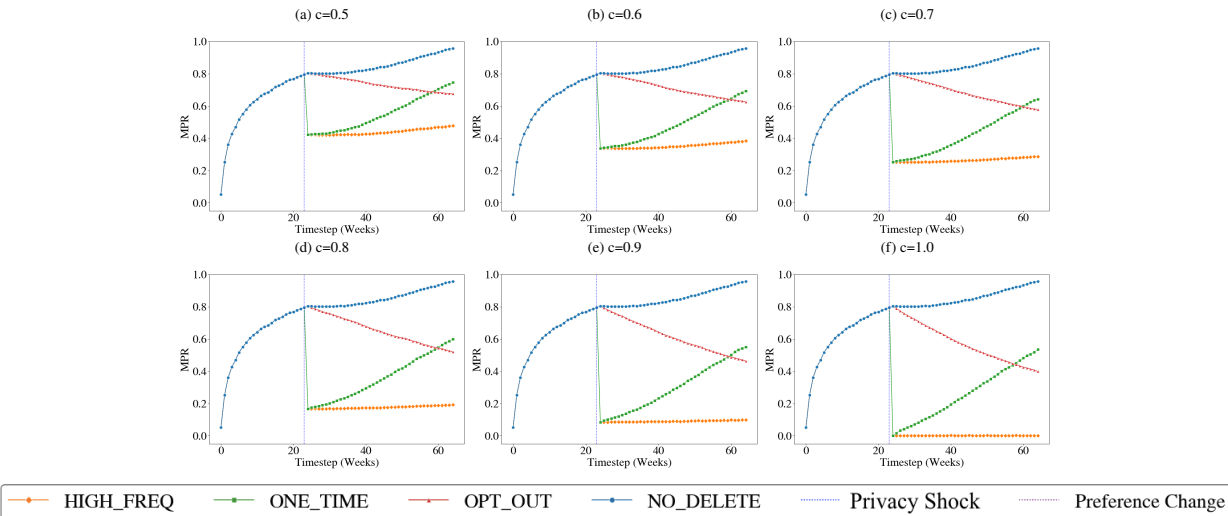
In Figures AA64 and AA65, we show longitudinal trends of the Mean Pairwise Revelation (MPR) metric for different values of niche threshold c under stable and dynamic user preference settings, respectively. The results remain highly consistent with Figure 9 in the main paper.

Figure AA64: MPR Trends for Data Privacy Practices with Varying Niche Thresholds, Stable User Preferences



Across all figures, 50% of users are affected, and the user reliance on RS is 20%. The algorithm is BPR. Setting $c = 1.0$ represents canonical, fully enacted (i.e., non-niche-based) data privacy practices.

Figure AA65: MPR Trends for Data Privacy Practices with Varying Niche Thresholds, Dynamic User Preferences



Across all figures, 50% of users are affected, the user reliance on RS is 20%, and the user preference change co-occurs with the privacy shock. The algorithm is BPR. Setting $c = 1.0$ represents canonical, fully enacted (i.e., non-niche-based) data privacy practices.